# Genetic dissection of the biotic stress response using a genome-scale gene network for rice

Insuk Lee[a,1,2], Young-Su Seo[b,1,3], Dusica Coltrane[b], Sohyun Hwang[a], Taeyun Oh[a], Edward M. Marcotte[c,d,2], and Pamela C. Ronald[b,e,f,2]

[a]Department of Biotechnology, College of Life Science and Biotechnology, Yonsei University, 262 Seongsanno, Seodaemun-gu, Seoul 120-749, Korea; [b]Department of Plant Pathology, University of California, Davis, CA 95616; [c]Center for Systems and Synthetic Biology, and [d]Department of Chemistry and Biochemistry, Institute for Cellular and Molecular Biology, University of Texas, Austin, TX 78712-1064; [e]Joint Bioenergy Institute, Emeryville, CA 94608; and [f]The Genome Center, University of California, Davis, California 95616

Edited by Steven P. Briggs, University of California at San Diego, La Jolla, CA, and approved October 11, 2011 (received for review June 28, 2011)

**Rice is a staple food for one-half the world's population and a model for other monocotyledonous species. Thus, efficient approaches for identifying key genes controlling simple or complex traits in rice have important biological, agricultural, and economic consequences. Here, we report on the construction of RiceNet, an experimentally tested genome-scale gene network for a monocotyledonous species. Many different datasets, derived from five different organisms including plants, animals, yeast, and humans, were evaluated, and 24 of the most useful were integrated into a statistical framework that allowed for the prediction of functional linkages between pairs of genes. Genes could be linked to traits by using guilt-by-association, predicting gene attributes on the basis of network neighbors. We applied RiceNet to an important agronomic trait, the biotic stress response. Using network guilt-by-association followed by focused protein–protein interaction assays, we identified and validated, in planta, two positive regulators, LOC_Os01g70580 (now Regulator of XA21; ROX1) and LOC_Os02g21510 (ROX2), and one negative regulator, LOC_Os06g12530 (ROX3). These proteins control resistance mediated by rice XA21, a pattern recognition receptor. We also showed that RiceNet can accurately predict gene function in another major monocotyledonous crop species, maize. RiceNet thus enables the identification of genes regulating important crop traits, facilitating engineering of pathways critical to crop productivity.**

systems biology | plant genetics | gene-trait associations

Rice (*Oryza sativa*) is the most important staple food crop. As one of the best studied grasses, rice also has an accumulated wealth of knowledge that makes it an attractive candidate as a reference for other important staple crops and emerging biofuel grasses (1). Its compact genome size (≈430 Mb), well-established methods for genetic transformation (2), availability of high-density genetic maps and whole-genome microarrays (reviewed in ref. 3), finished genome sequence (4), and close relationships with other cereals, all make rice an ideal model system in which to study plant physiology, development, agronomics, and genomics of grasses (5–7). Furthermore, in recent years, several laboratories have successfully developed gene-indexed mutants for targeted loss-of-function or gain-of-function analysis of many rice genes (reviewed in ref. 3).

These advances have led to the accumulation of sufficient public data to construct systems-level models of rice gene interactions. In principle, such models should allow for the prediction and systematic discovery of genes and associated pathways that control phenotypes of economic importance, such as tolerance to environmental stress and resistance to disease. We have developed such a network modeling platform, called a probabilistic functional gene network (8, 9), variations of which have been successfully applied to predict novel gene functions in e.g., yeast (10), worm (11), mice (12, 13), humans (14–17), and *Arabidopsis thaliana* (18). In principle, *Arabidopsis* gene networks should be useful even for rice genes, as many processes are conserved between dicotyledonous species, including *Arabidopsis*, and monocotyledonous species, including rice. However, monocots and dicots diverged >160–200 million years ago; thus, many gene networks differ significantly between these two main groups of flowering plants (19–25). Therefore, a complete understanding of monocot gene networks will depend on a full characterization of these pathways in an experimentally tractable monocotyledonous species such as rice. Here, we present an experimentally validated genome-scale functional gene network of a monocotyledonous species, a network of rice genes, named RiceNet, reconstructed from quantitative integration of available genomics and proteomics datasets.

Construction of a genome-wide network for rice is challenging for several reasons. First, whereas *A. thaliana* has ≈27,000 protein coding genes (The *Arabidopsis* Information Resource, release 9; ref. 26), rice has 41,203 nontransposable element (TE)-related protein coding genes [The Institute for Genomic Research (TIGR) rice annotation release 5; ref. 27]. This increased genome complexity results in a combinatorial explosion for the number of hypotheses for pairwise relations between genes, complicating discovery of true functional associations. Second, the current reference knowledge and raw genomic data available for models are much sparser for rice than for *Arabidopsis*, reducing predictive power of models. Third, the experimental validation of predicted gene function is more difficult in rice than *Arabidopsis* because of a longer reproductive cycle, larger plant sizes, greenhouse requirements, fewer available gene knockout strains, and less efficient transformation procedures (3). Despite these hurdles, we reconstructed a network covering ≈50% of the 41,203 rice genes. This network builds on a published midsized network of 100 rice stress response proteins that was constructed through protein interaction mapping (28). We demonstrated that RiceNet associations are highly predictive for diverse biological processes in rice. We further predicted and experimentally validated three previously unknown regulators of resistance mediated by XA21, a rice pattern recognition receptor that is a key determinant of the innate immune response (29). RiceNet also showed significant predictive power for identifying genes that function in the stress response of another major crop, maize. These results indicate that RiceNet can accurately predict gene function in monocotyledonous species.

Author contributions: I.L., Y.-S.S., E.M.M., and P.C.R. designed research; I.L., Y.-S.S., D.C., E.M.M., and P.C.R. performed research; I.L., Y.-S.S., S.H., T.O., E.M.M., and P.C.R. analyzed data; and I.L., E.M.M., and P.C.R. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]I.L. and Y.-S.S. contributed equally to this work.

[2]To whom correspondence may be addressed. E-mail: insuklee@yonsei.ac.kr, marcotte@icmb.utexas.edu, or pcronald@ucdavis.edu.

[3]Present address: Department of Microbiology, Pusan National University, Busan 609-735, Republic of Korea.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1110384108/-/DCSupplemental.

## Results

**RiceNet: A Genome-Scale Gene Network for a Monocotyledonous Species.** We aimed to construct a gene network of rice spanning as many as feasible of the 41,203 non-TE–related protein coding genes annotated by the TIGR Rice Genome Annotation Release 5 (27). Only limited numbers of genome-scale datasets are available for rice. Such limitations, in turn, reduce the scale of networks that can be reconstructed (e.g., a rice gene network based on only mRNA coexpression in all available rice gene microarrays covered only ≈10% of the genome; ref. 30). This shortcoming can be partially overcome by transferring datasets from other organisms via the use of gene orthology relationships (31). The key to using such data lies in the judicious weighting of datasets from other species to maximize reconstruction of conserved rice gene systems, while not degrading reconstruction of rice-specific gene systems from datasets collected in rice. The value of this approach has been shown in reconstruction of gene networks for *Caenorhabditis elegans* (11) and *Arabidopsis* (18). Thus, in addition to rice (*Oryza sativa*) datasets, we also identified evolutionarily conserved gene–gene linkages between rice genes by using datasets from *Saccharomyces cerevisiae*, *C. elegans*, *Homo sapiens*, and *A. thaliana*.

A total of 24 different types of data (spanning many individual datasets) were quantitatively integrated into a single gene network as described in full in the *SI Appendix*. The datasets used to infer functional linkages spanned many types of gene–gene relationships, including both direct measurements of physical and genetic interactions, as well as inferred interactions from genome sequences, literature mining, and protein structures. In all, the 24 data types, from five different organisms, included transcript coexpression links based on several hundred DNA microarray datasets (e.g., *SI Appendix,* Table S1 for rice), genome-scale protein–protein physical interactions mapped by yeast two hybrid or affinity purification followed by mass spectrometry identification of protein complexes, linkages between genes automatically mined from PubMed articles, protein–protein interactions from curated databases, linkages between proteins with similar domain co-occurrence profiles, genetic interactions, linkages based on genes' phylogenetic profiles or tendencies for bacterial orthologs to occur as neighbors in many bacterial genomes, and protein–protein interactions inferred from protein tertiary structures (Fig. 1*A* and *SI Appendix,* Table S2). Clearly, differing lines of evidence provide widely differing degrees of support for functional linkages. For each pair of rice genes, the evidence from all datasets was integrated into a single numerical log likelihood score (LLS) denoting the likelihood for those genes to function together as supported by sharing Gene Ontology biological process (GO-BP) terms annotated by the TIGR Rice Genome Annotation Release 5 (27) (*SI Appendix*). Integration improved genome coverage and linkage accuracy beyond all of the individual datasets (Fig. 1*A*). The final, integrated set of gene–gene linkages, named RiceNet (www.functionalnet.org/ricenet), contains a total of 588,221 links connecting 18,377 non-TE–related rice protein-coding genes (44.6% of 41,203 genes) (*SI Appendix,* Fig. S1). RiceNet also covers 16,678 (63.7%) of the 26,178 rice genes known, thus far, to be expressed.

**RiceNet Is More Accurate and Extensive than a Network Generated by Orthology from *Arabidopsis*.** It is an open question how well gene networks derived from better-characterized dicots such as *Arabidopsis* might faithfully reconstruct pathways and systems in a monocot. For example, an alternate approach to constructing a rice gene network might be simply to transfer linkages from orthologous gene pairs from the existing *Arabidopsis* gene network, AraNet (18). This approach does not require modeling using rice annotations or any of the rice-derived experimental data. To assess the accuracy of such a network, we first defined an AraNet-derived network with the same number of functional links as those of a RiceNet. The AraNet-derived network covers 12,225 rice genes with 588,221 links, whereas RiceNet covers 18,377 genes (6,122 more genes) with the same number of links (Fig. 1*B*). We tested the accuracy of the AraNet-derived network versus RiceNet using linkages from the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database (32), which is based on manual curation. KEGG is thus considered generally accurate and largely independent from both RiceNet and AraNet, and shares only 2.2% of the 662,936 GO-BP–derived linkages used to guide the integration of RiceNet (*SI Appendix*). To err on the conservative side, we excluded KEGG linkages shared with GO-BP so as to obtain a set of 89,140 KEGG linkages that were fully independent from the linkages used to guide RiceNet construction. As expected, gene links supported by both networks are more accurate than network links predicted by only one approach. However, RiceNet-specific linkages are ≈2.5 times more accurate than those derived solely by orthology from AraNet (15.5% versus 6.5% true positive rate; Fig. 1*B*). In terms of genome coverage, RiceNet covers 4,839 additional genes (≈12% of rice genome). Thus, reconstructing a gene network specifically for rice genes improves both accuracy and coverage of the network.

**RiceNet Reflects Well-Defined Biological Pathways and Processes.** We assessed the quality of RiceNet for modeling biological processes by several additional computational analyses. First, we used topological analysis to assess whether the RiceNet contains modular structures consistent with well-defined biological pathways and processes. We found that RiceNet shows a 100-fold higher clustering coefficient (33) than a randomized network (*SI Appendix,* Fig. S2*A*). The observed higher extent of clustering is an expected characteristic of functional modules. Similarly, we observed very nonrandom path lengths connecting gene pairs in RiceNet (*SI Appendix,* Fig. S2*B*), indicating tightly interconnected regional structures (representing functional modules) separated by longer chains of functional associations. Both topological properties suggest RiceNet is organized into gene modules separated in the network.

Second, we determined whether the RiceNet-predicted gene modules reflect known biological pathways in rice. In this "guilt-by-association" approach, we prioritized candidate genes for each biological process based on network connections to known genes in those processes, assessing predictive accuracy by using cross-validation and receiver operating characteristic (ROC) analysis. Our previous study demonstrated that superior prediction performance can be achieved by using methods that consider not only direct network neighbors but also indirect ones (34). Therefore, we prioritized candidate genes by using both direct and indirect network neighbors via Gaussian smoothing (17). Prediction power can be summarized from a ROC analysis as the area under the ROC curve (AUC), which ranges from near 0.5 for random expectation to 1 for perfect predictions. For a total of 834 GO-BPs with more than three annotated genes, 642 processes (77% of total tested GO-BPs) showed AUC > 0.7, which indicates that RiceNet is highly predictive of gene function (Fig. 1*C*), far in excess of chance expectation [e.g., randomized gene sets of the same sizes show a median AUC of 0.5 and only 25 terms (3%) show AUC > 0.7]. These results strongly suggest that RiceNet is predictive for diverse types of biological pathways in rice. Moreover, RiceNet proved far more predictive of rice gene function than AraNet, as tested for GO-BP terms ($P < 1 \times 10^{-65}$; Wilcoxon signed rank sum test).

Third, we examined cell type-specific mRNA expression of rice genes. Rice, like other multicellular organisms, is composed of many distinct cell types. Each cell type expresses a unique set of genes that together produce the characteristics of that cell type. In contrast, RiceNet is composed of just one integrated network spanning a large fraction of the set of rice genes. Although cell-type specificity is not explicitly modeled in RiceNet,

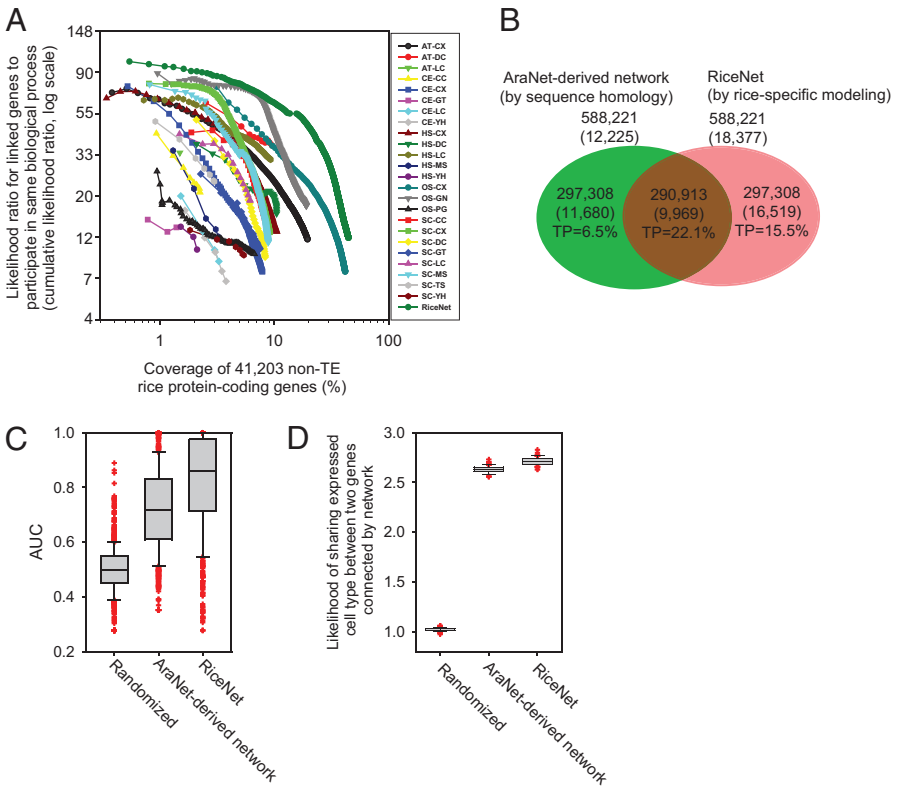



**Fig. 1.** Summary of construction and computational assessment of RiceNet, a genome-scale functional gene network for *O. sativa*. (*A*) Pairwise gene linkages derived from 24 diverse functional genomics and proteomics data types, each spanning many individual experiments and representing in all >60 million experimental or computational observations, were integrated into a composite gene network with higher accuracy and genome coverage than any individual dataset. The integrated network (RiceNet) contains 588,221 functional linkages among 18,377 (45%) of the 41,203 non-TE–related protein-coding rice genes. The plot *x* axis indicates the log-scale percentage of the 41,203 protein-coding genes covered by functional linkages derived from the indicated datasets (plotted curves); the *y* axis indicates the accuracy of functional linkages derived from the datasets, measured as the cumulative likelihood for linked genes to share GO-BP term annotations, tested using 0.632 bootstrapping and plotted for successive bins of 1,000 linkages each (symbols). Datasets are named as XX-YY, where XX indicates species of data origin (AT, *A. thaliana*; CE, *C. elegans*; HS, *H. sapiens*; OS, *O. Sativa;* SC, *S. cerevisiae*) and YY indicates data type (CC, cocitation; CX, mRNA coexpression; DC, domain co-occurrence; GN, gene neighbor; GT, genetic interaction; LC, literature curated protein interactions; MS, affinity purification/mass spectrometry; PG, phylogenetic profiles; TS, tertiary structure; YH, yeast two hybrid). (*B*) RiceNet includes many linkages beyond those found by simple orthology from the *Arabidopsis* gene network AraNet (18), as shown by a Venn diagram of the gene linkages. RiceNet covers many more rice genes (18,377 versus 12,225 genes) than the AraNet-derived network with the same number of links. Measurement of linkage accuracy show that links supported by both networks are more accurate (true positive rate; TP = 22.1%) than those by only one network. The higher accuracy of RiceNet-only links (TP = 15.5%) than that of AraNet-transferred links (TP = 6.5%) indicates the value of optimizing the functional network for rice. Linkage accuracy was measured by using a completely independent set of reference linkages derived from the KEGG biological pathway database only. (*C*) Pathway predictability is generally high, as measured by the areas under cross-validated ROC curves for correctly prioritizing known genes for 834 GO-BP annotations (with more than three associated genes) using network guilt-by-association via Gaussian smoothing (17). In bar-and-whiskers plots, the central horizontal line in the box indicates the median AUC, the boundaries of the box indicate the first and third quartiles, whiskers indicate the 10th and 90th percentiles of the AUC distribution, and plus signs indicate outliers. RiceNet is superior to the AraNet-derived network and significantly outperforms randomized tests. (*D*) RiceNet links genes with similar cell-type specific expression patterns. Genes connected by RiceNet were significantly more coexpressed (by nearly threefold) across 40 individual rice cell types (35) than were genes linked in randomized networks (repeating the calculation for 100 randomized networks and plotting the distribution of the 100 resulting odds ratios), and somewhat more than for linkages transferred by orthology from AraNet (*P* value $< 1 \times 10^{-17}$; Wilcoxon signed rank sum test).

we tested whether it nonetheless can be used to model cell-type specific functions. Assuming that many biological processes are carried out through functionally specialized cell types, functional associations are expected to be enriched among genes expressed in the same cell-types. We measured the likelihood of genes connected in RiceNet or the AraNet-derived network to share cell-type specific expression (*SI Appendix*). For this analysis, we used the rice transcriptome atlas database, which profiles transcript expression across 40 rice cell types (35). We found that RiceNet links genes expressed in the same cell-types at nearly three times the rate of a randomized network (Fig. 1*D*). AraNet-derived links show somewhat lower enrichment for cell-type specificity ($P < 1 \times 10^{-17}$; Wilcoxon signed rank sum test), indicating that optimizing the network for rice contributed to cell-type specificity, and may help explain the higher accuracy of RiceNet (Fig. 1*B*).

**Two-Step Network-Guided Discovery of ROX1, ROX2, and ROX3, Three Regulators of XA21-Mediated Immunity.** Our pathway analysis described above demonstrates that genes for similar biological processes can be successfully associated in RiceNet. We next specifically tested the feasibility of identifying previously unknown genes governing biotic stress response pathways by using RiceNet. We reported the construction of a stress response interactome consisting of 100 proteins; 46 of these proteins are predicted to be involved in the biotic stress response (28). Fifteen of these interactome components are of particular interest because they have been confirmed to play a key role in the rice defense response by using loss-of-function and gain-of-function analyses (28). We therefore used these 15 genes to query RiceNet in an attempt to identify novel candidate genes governing XA21-mediated immunity (Fig. 2 and *SI Appendix*, Table S3). Note that *Xa21* itself is not present in RiceNet, because it is not present in

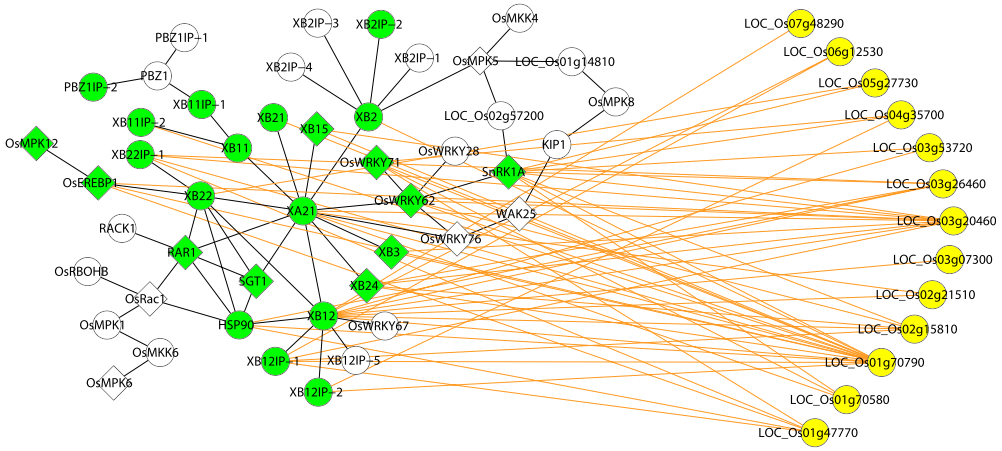



**Fig. 2.** Extensive interactions among RiceNet-predicted proteins with other proteins predicted to be involved in XA21-mediated immunity. Thirteen of the 14 candidate genes predicted by RiceNet (*SI Appendix,* Table S3) to function in XA21-mediated immunity are indicated by yellow circles. Pairwise protein–protein interactions between these 13 proteins and 24 components (green nodes) of the previously reported biotic stress response subinteractome (28) are designated by orange edges. The 14th RiceNet-predicted protein is not shown because it did not bind with any of the 24 tested interactome members using yeast two-hybrid assays as indicated in the text. Other members of the interactome (white nodes) and their connections (black edges) are included for context. Diamonds indicate the 15 proteins validated phenotypically, using loss-of-function and gain-of-function analyses, to modulate rice defense responses. These 15 proteins served as the RiceNet query as described in the text. All constructs carry full-length cDNAs except for the *Xa21* construct as previously described (36).

the *Oryza sativa* genome (it was isolated from *O. longistaminata)* (36). For this reason, *Xa21* was not used as part of the query set.

When the 15 genes with validated phenotypes were used to query RiceNet, we identified 802 rice genes with a predicted function in XA21-mediated immunity. To further prioritize within this set, we selected only genes connected to ≥1 query gene by LLS ≥2.0, connected to ≥2 query genes by LLS ≥1.5, or that had not been demonstrated to confer a function in plant immune responses based on literature searches (*SI Appendix,* Table S4). These criteria narrowed the list to 14 genes. We then experimentally assayed the interactions of the proteins encoded by these genes with an available set of 24 proteins from the biotic stress response subinteractome, including XA21 itself (28) (*SI Appendix,* Table S3 and *SI Appendix,* Table S5). Thirteen of the 14 candidate genes interacted by yeast two hybrid assays with at least one component of the XA21 interactome (Fig. 2), confirming the value of this prioritization strategy.

We selected five of these candidates for characterization in planta based either on their direct interactions with XA21 (LOC_Os01g70580, LOC_Os01g70790, LOC_Os02g21510, and LOC_Os03g20460) or on their interactions with proteins harboring sequence motifs related to the inflammatory response in animals (LOC_Os06g12530) (*SI Appendix,* Table S4). We generated overexpression (ox) and RNAi constructs for each gene (except for LOC_Os06g12530 for which only an overexpression construct was generated) and introduced these constructs into a homozygous Kitaake-XA21 rice line using a hygromycin selectable marker. We then assayed for resistance to the bacterial pathogen *Xanthomonas oryzae pv. oryzae* (*Xoo*), measuring lengths of water-soaked lesions 14–21 d after inoculation.

We did not observe any obvious differences between the control Kitaake-XA21 and transgenic lines either overexpressing or silenced for LOC_Os01g70790 or LOC_Os03g20460 (*SI Appendix,* Table S6). However, transgenic plants silenced for two of the genes (LOC_Os01g70580 RNAi, LOC_Os02g21510 RNAi) showed clear enhanced susceptibility to *Xoo* as compared with the Kitaake-XA21 control. Transgenic plants overexpressing LOC_Os06g12530 showed enhanced susceptibility to *Xoo* compared with the Kitaake-XA21 control (*SI Appendix,* Fig. S3 and Fig. 3). These genes were designated *Rox1* (Regulator of XA21-mediated immunity 1), *Rox2*, and *Rox3*, respectively. The phenotypes are heritable for two generations, the RNA levels correlate with the overexpression or knockdown, and the transgene cosegregates with the altered phenotypes in progeny analyses (*SI Appendix*). Based on the phenotypes, *Rox1* and Rox2 are positive regulators (silencing leads to enhanced susceptibility) and *Rox3* is a negative regulator (overexpression leads to enhanced susceptibility). Thus, we were able to validate three of five RiceNet predictions in planta (60% success rate).

Sequence homology sheds some light on more specific roles for these three ROX proteins. First, the positive regulator *Rox1* is annotated as a thiamine pyrophosphokinase (TPK). TPK catalyzes the transfer of a pyrophosphate group from ATP to vitamin B1 (thiamine) to form the coenzyme thiamine pyrophosphate (TPP). It has been reported that treatment of thiamine and TPP induces resistance to rice pathogens including *Xoo* (37). These results further support a role for *Rox1* in the rice defense responses.

The positive regulator *Rox2* is a member of the NOL1/NOL2/sun gene family. Human homologs have been implicated in Williams-Beuren syndrome, a developmental disorder associated with haploinsufficiency of multiple genes at 7q11.23 (38). NOL1/NOL2/sun gene family members have not been shown to function in innate immunity.

Finally, the negative regulator *Rox3* is annotated as a nuclear migration protein, nudC. nudC plays a key role in cell division through the regulation of cytoplasmic dynein and also in the regulation of the inflammatory response (39). These results indicate that sequence homology alone would not have suggested roles in immunity, emphasizing the value of RiceNet for identifying new genes relevant to the biotic response in the absence of strong a priori knowledge.

**RiceNet Predictability Extends to Another Monocotyledonous Crop Species, Maize.** Given that RiceNet efficiently predicts gene function in rice, we asked whether this predictability extends to another monocotyledonous crop species, maize (*Zea mays*). Because RiceNet is more accurate for rice gene function than a network generated by orthology from *Arabidopsis* (Fig. 1*B*), we also hypothesized that RiceNet would be more efficient than AraNet at gene prediction in another monocotyledonous species. We therefore compared the predictive power of an AraNet-derived maize gene network (AT-ZM), a RiceNet-derived maize gene network (OS-ZM), and maize GO-BP annotation by AgBase (40) (*SI Appendix*).

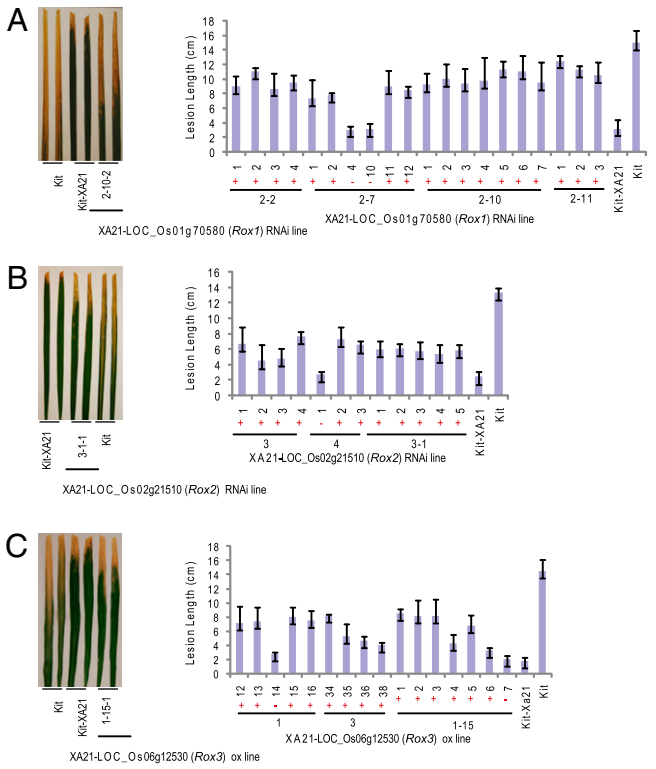



**Fig. 3.** Three new regulators of XA21 (ROX)-mediated immunity predicted by RiceNet are validated by *Xoo* infection assays of transgenic rice plants. The response to *Xoo* was assayed in infected leaves of 5-wk-old rice plants. A symbol (-) indicates lack of the transgene, and a symbol (+) indicates the presence of the transgene. (*A*) LOC_Os01g70580 (ROX1) is a positive regulator of XA21-mediated immunity. Kitaake-XA21 (Kit-XA21) plants with RNAi-mediated suppression of LOC_Os01g70580 (*Rox1*) were generated and assayed for resistance. Lesion lengths were measured in leaves of 5-wk-old rice plants from T2 progeny of the XA21- LOC_Os01g70580 (ROX1) RNAi 2-2, 2-7, 2-10, and 2-11 lines 14 d after *Xoo* inoculation. (*B*) LOC_Os02g21510 (ROX2) is a second positive regulator of XA21-mediated immunity. Kitaake-XA21 plants with RNAi-mediated suppression of LOC_Os02g21510 (*Rox2*) were generated and lesion lengths were measured as above in leaves of T1 progeny from XA21-LOC_Os02g21510 (*Rox2*) RNAi 3 and 4 lines and T2 progeny from the XA21-LOC_Os02g21510 (*Rox2*) RNAi 3-1 line. (*C*) LOC_Os06g12530 (*Rox3*) is a negative regulator of XA21-mediated immunity. Kitaake-XA21 plants overexpressing (ox) LOC_Os06g12530 (*Rox3*) were generated and response to *Xoo* was measured as above by using leaves from T1 progeny from the LOC_Os06g12530 (*Rox3*) overexpression lines, 1 and 3, and T2 progeny from the LOC_Os06g12530 (*Rox3*) overexpression 1-15 line. Each bar represents the average and SD from at least three tested leaves. Kitaake-XA21 and Kitaake (Kit) were used as controls. Photographs of lesion length were taken at 14 d after *Xoo* inoculation. Additional genetic tests and expression quantification of candidate genes in transgenic lines are described in *SI Appendix* and *SI Appendix,* Fig. S3.

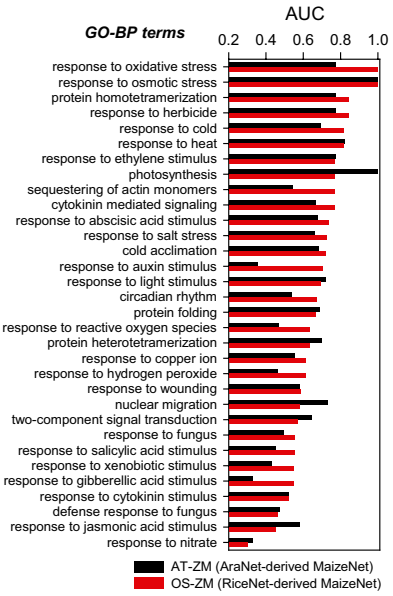



**Fig. 4.** RiceNet predicts maize biological processes. Thirty-two maize GO-BP gene sets, each with three or more associated genes identified through noncomputational approaches, were tested for predictability by the AraNet-derived and RiceNet-derived maize gene networks, indicated by AT-ZM and OS-ZM, respectively. The predictability of each process was measured by using cross-validation as the area under a ROC curve (AUC) with Gaussian smoothing methods (17). A total of 30 maize GO-BP terms are predictable (AUC > 0.5) by either AT-ZM or OS-ZM. For 20 of the 30, OS-ZM outperformed AT-ZM. The distribution of AUC scores by OS-ZM was significantly higher than that by AT-ZM ($P$ value = $1.42 \times 10^{-2}$; Wilcoxon signed rank sum test).

RiceNet appears predictive for many maize gene pathways. We identified 32 GO-BP terms that have three or more annotated maize genes based on experimental or literature evidence and used these 32 GO-BP terms with highly reliable annotations for the analysis. Using cross-validated ROC analysis as in ref. 41, we tested the two networks (OS-ZM and AT-ZM), for correctly prioritizing genes in these GO-BPs, using the Gaussian smoothing guilt-by-association method (17). Both networks showed good predictability for maize GO-BP terms (Fig. 4 and *SI Appendix,* Table S7). For 30 terms predictable (AUC > 0.5) by either network, OS-ZM performed significantly better than AT-ZM ($P$ value = $1.42 \times 10^{-2}$; Wilcoxon signed rank sum test). This predictability was independent of the number of associated genes (*SI Appendix,* Fig. S4). Noticeably, many abiotic stress responses in maize are highly predictable with OS-ZM. Thus, RiceNet is useful for studies of gene-trait associations in both rice and maize, and this work suggests that RiceNet will also be useful for predicting gene function in other important monocotyledonous species such as wheat and switchgrass, for which species-specific networks have not yet been constructed.

## Discussion

Here, we report construction of RiceNet, a genome-scale gene network for rice. We demonstrate its predictive power for diverse biological processes and its usefulness in identifying genes governing rice innate immunity. RiceNet represents an experimentally validated genome-scale gene network for a crop species. Using RiceNet, we systematically selected candidate genes predicted to be involved in the rice defense response by using a two-step network-guided prediction approach—first by guilt-by-association, and then by protein–protein interaction tests. From rough computational estimates, we expect RiceNet to often offer correct candidate genes when the predictive AUC is high (e.g., >0.75), with the number of correct candidates increasing as AUC increases (*SI Appendix,* Fig. S5). The additional prioritization offered by protein interaction screening of candidates appeared to elevate the validation rate even further, and in our transgenic tests of five of the candidate genes selected by using this two-step strategy, we confirmed three genes as regulators of XA21-mediated immunity. Given the large genomes of most crop species (generally 30,000–50,000 genes) and their long reproductive cycles (often several months), this two-step network-guided prioritization should facilitate identification of key genes controlling important traits and future engineering of agronomically useful varieties.

Several other features of RiceNet are particularly notable, including that a network optimized for rice outperformed an alternative rice gene network constructed by orthology from the dicot *Arabidopsis* (18) (Fig. 1 *B* and *D*). We expect that other

crop species will have similar requirements, benefiting from species-specific datasets and optimization. Given the impressive advances in genomic and proteomic technology for food crop species (42, 43) and candidate bioenergy crops (44) and other models (45), a sharp rise in available datasets appears likely for these species over the next few years. For rice, additional rice-specific experimental data including protein–protein interactions and gene expression data spanning new experimental conditions appear most valuable for improving the accuracy and coverage of RiceNet into the future.

We have shown that RiceNet is useful for predicting gene function in maize, another monocotyledonous crop species. The high predictability for maize genes by a RiceNet-derived maize gene network (OS-ZM) and its superior performance compared with an AraNet-derived maize gene network (AT-ZM) indicates that monocot-specific modeling improves gene prediction for other monocotyledonous crop species. This finding is particularly important in light of the lack of species-specific gene networks for maize and other monocotyledonous crop species.

Economically important crop traits range from simple traits emerging from strong selective pressure during domestication to highly genetically complex traits (46). Manipulation of only a few key genes related to the traits may not generate the desired phenotypes. Therefore, approaches for broadly defining relevant genes, coupled with rapid and inexpensive interaction assays to more finely prioritize genes, offer an attractive and potentially rapid route for focusing crop engineering efforts on the small sets of genes that are deemed most likely to affect the traits of interest.

## Methods

Reference and benchmark sets, raw datasets, and computational methods of construction and analysis of RiceNet are described in full in *SI Appendix*.

Construction of MaizeNet by orthology-based links from AraNet and RiceNet are described in full in *SI Appendix*.

Genetic analysis of ROX1, ROX2, and ROX3 transgenic plants are described in full in *SI Appendix*.

A user-interactive web tool for RiceNet-based selection of candidate genes is publicly available at http://www.functionalnet.org/ricenet.

**ACKNOWLEDGMENTS.** This work was supported by the US Department of Energy, Office of Science, Office of Biological and Environmental Research, through Contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the US Department of Energy. This work was also supported by the National Research Foundation of Korea funded by the Korean government Ministry of Education, Science, and Technology Grants 2010-0017649 and 2010-0001818 and POSCO TJ Park Science fellowship (to I.L.); the National Science Foundation, National Institutes of Health (NIH), Welch Foundation Grant F1515, and Packard Foundation (to E.M.M.); and NIH Grant GM 55962 (to P.C.R.).

1. Paterson AH, Bowers JE, Peterson DG, Estill JC, Chapman BA (2003) Structure and evolution of cereal genomes. *Curr Opin Genet Dev* 13:644–650.
2. Hiei Y, Ohta S, Komari T, Kumashiro T (1994) Efficient transformation of rice (Oryza sativa L.) mediated by Agrobacterium and sequence analysis of the boundaries of the T-DNA. *Plant J* 6:271–282.
3. 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.
4. Matsumoto T, et al.; International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. *Nature* 436:793–800.
5. Gale MD, Devos KM (1998) Comparative genetics in the grasses. *Proc Natl Acad Sci USA* 95:1971–1974.
6. Goff SA (1999) Rice as a model for cereal genomics. *Curr Opin Plant Biol* 2:86–89.
7. Shimamoto K, Kyozuka J (2002) Rice as a model for comparative genomics of plants. *Annu Rev Plant Biol* 53:399–419.
8. Lee I (2011) Probabilistic functional gene societies. *Prog Biophys Mol Biol* 106:435–442.
9. Lee I, Date SV, Adai AT, Marcotte EM (2004) A probabilistic functional network of yeast genes. *Science* 306:1555–1558.
10. Li Z, et al. (2009) Rational extension of the ribosome biogenesis pathway using network-guided genetics. *PLoS Biol* 7:e1000213.
11. Lee I, et al. (2008) A single gene network accurately predicts phenotypic effects of gene perturbation in Caenorhabditis elegans. *Nat Genet* 40:181–188.
12. Kim WK, Krumpelman C, Marcotte EM (2008) Inferring mouse gene functions from genomic-scale data using a combined functional network/classification strategy. *Genome Biol* 9(Suppl 1):S5.
13. Peña-Castillo L, et al. (2008) A critical assessment of Mus musculus gene function prediction using integrated genomic evidence. *Genome Biol* 9(Suppl 1):S2.
14. Huttenhower C, et al. (2009) Exploring the human genome with functional maps. *Genome Res* 19:1093–1106.
15. Lee I, Blom UM, Wang PI, Shim JE, Marcotte EM (2011) Prioritizing candidate disease genes by network-based boosting of genome-wide association data. *Genome Res* 21:1109–1121.
16. Linghu B, Snitkin ES, Hu Z, Xia Y, Delisi C (2009) Genome-wide prioritization of disease genes and identification of disease-disease associations from an integrated human functional linkage network. *Genome Biol* 10:R91.
17. Mostafavi S, Ray D, Warde-Farley D, Grouios C, Morris Q (2008) GeneMANIA: A real-time multiple association network integration algorithm for predicting gene function. *Genome Biol* 9(Suppl 1):S4.
18. Lee I, Ambaru B, Thakkar P, Marcotte EM, Rhee SY (2010) Rational association of genes with traits using a genome-scale gene network for Arabidopsis thaliana. *Nat Biotechnol* 28:149–156.
19. Flavell R (2009) Role of model plant species. *Methods Mol Biol* 513:1–18.
20. Chern M, Canlas PE, Fitzgerald HA, Ronald PC (2005) Rice NRR, a negative regulator of disease resistance, interacts with Arabidopsis NPR1 and rice NH1. *Plant J* 43:623–635.
21. 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.
22. Devos KM, Beales J, Nagamura Y, Sasaki T (1999) Arabidopsis-rice: Will colinearity allow gene prediction across the eudicot-monocot divide? *Genome Res* 9:825–829.
23. Fitzgerald HA, Canlas PE, Chern MS, Ronald PC (2005) Alteration of TGA factor activity in rice results in enhanced tolerance to Xanthomonas oryzae pv. oryzae. *Plant J* 43: 335–347.
24. Fitzgerald HA, Chern MS, Navarre R, Ronald PC (2004) Overexpression of (At)NPR1 in rice leads to a BTH- and environment-induced lesion-mimic/cell death phenotype. *Mol Plant Microbe Interact* 17:140–151.
25. Xu K, et al. (2006) Sub1A is an ethylene-response-factor-like gene that confers submergence tolerance to rice. *Nature* 442:705–708.
26. Swarbreck D, et al. (2008) The Arabidopsis Information Resource (TAIR): Gene structure and function annotation. *Nucleic Acids Res* 36(Database issue):D1009–D1014.
27. Ouyang S, et al. (2007) The TIGR Rice Genome Annotation Resource: Improvements and new features. *Nucleic Acids Res* 35(Database issue):D883–D887.
28. Seo YS, et al. (2011) Towards establishment of a rice stress response interactome. *PLoS Genet* 7:e1002020.
29. Ronald PC, Beutler B (2010) Plant and animal sensors of conserved microbial signatures. *Science* 330:1061–1064.
30. Ficklin SP, Luo F, Feltus FA (2010) The association of multiple interacting genes with specific phenotypes in rice using gene coexpression networks. *Plant Physiol* 154:13–24.
31. Remm M, Storm CE, Sonnhammer EL (2001) Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *J Mol Biol* 314:1041–1052.
32. Kanehisa M, Goto S, Kawashima S, Nakaya A (2002) The KEGG databases at GenomeNet. *Nucleic Acids Res* 30:42–46.
33. Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. *Nature* 393:440–442.
34. Wang PI, Marcotte EM (2010) It's the machine that matters: Predicting gene function and phenotype from protein networks. *J Proteomics* 73:2277–2289.
35. Jiao Y, et al. (2009) A transcriptome atlas of rice cell types uncovers cellular, functional and developmental hierarchies. *Nat Genet* 41:258–263.
36. Song WY, et al. (1995) A receptor kinase-like protein encoded by the rice disease resistance gene, Xa21. *Science* 270:1804–1806.
37. Ahn IP, Kim S, Lee YH (2005) Vitamin B1 functions as an activator of plant disease resistance. *Plant Physiol* 138:1505–1515.
38. Merla G, Ucla C, Guipponi M, Reymond A (2002) Identification of additional transcripts in the Williams-Beuren syndrome critical region. *Hum Genet* 110:429–438.
39. Riera J, Lazo PS (2009) The mammalian NudC-like genes: A family with functions other than regulating nuclear distribution. *Cell Mol Life Sci* 66:2383–2390.
40. McCarthy FM, et al. (2011) AgBase: Supporting functional modeling in agricultural organisms. *Nucleic Acids Res* 39(Database issue):D497–D506.
41. Lehner B, Lee I (2008) Network-guided genetic screening: Building, testing and using gene networks to predict gene function. *Brief Funct Genomics Proteomics* 7:217–227.
42. Schnable PS, et al. (2009) The B73 maize genome: Complexity, diversity, and dynamics. *Science* 326:1112–1115.
43. Schmutz J, et al. (2010) Genome sequence of the palaeopolyploid soybean. *Nature* 463:178–183.
44. Yuan JS, Tiller KH, Al-Ahmad H, Stewart NR, Stewart CN, Jr. (2008) Plants to power: Bioenergy to fuel the future. *Trends Plant Sci* 13:421–429.
45. Vogel JP, et al.; International Brachypodium Initiative (2010) Genome sequencing and analysis of the model grass Brachypodium distachyon. *Nature* 463:763–768.
46. Holland JB (2007) Genetic architecture of complex traits in plants. *Curr Opin Plant Biol* 10:156–161.

# Supplementary Information

**Genetic dissection of the biotic stress response using a genome-scale gene network for rice**

Insuk Lee, Young-Su Seo, Dusica Coltrane, Sohyun Hwang, Taeyun Oh, Edward M. Marcotte, and Pamela C. Ronald

**Overview of network construction**

RiceNet was constructed using the approach previously developed for *Arabidopsis* genes (1), customized for the particular demands of rice datasets, which include sparse rice-specific data and a lower proportion of genes of known function. Briefly, we collected diverse functional genomics, proteomics, and comparative genomics datasets likely to be relevant to rice biological processes. We then scored individual datasets for their ability to reconstruct known functional gene-gene associations in rice (using rice Gene Ontology Biological Processes as our reference set of known associations, as described in more detail below) using the Bayesian log likelihood scoring scheme as described for AraNet (1). Within this framework, we estimated the likelihood of two genes participating in the same process conditioned on each dataset using a log likelihood score (LLS) (described in (2)), and then combined scored linkages derived from the various datasets to construct an integrated gene network.

In this scheme, $\mathrm{LLS} = \ln\left(\frac{\mathrm{P(L \mid E)}/\mathrm{P(\neg L \mid E)}}{\mathrm{P(L)}/\mathrm{P(\neg L)}}\right)$, where P(L|E) and P(¬L|E) are the frequencies of linkages (L) observed in experiment (E) between annotated genes operating in the same biological processes and in different biological processes, respectively, while P(L) and P(¬L) represent the prior expectations (*i.e.*, the total frequency of linkages between all annotated rice genes operating in the same biological processes and operating in different biological processes, respectively). Higher scores indicate stronger support for the genes operating in the same biological processes. We monitored overtraining of the network model by using 0.632 bootstrapping for all LLS evaluations as in (1). Following this initial scoring, we integrated linkages from different datasets and different lines of evidence into a single network, which now incorporated diverse lines of evidence into its linkages. For this integration, we employed the weighted sum method as described in (3). The resulting network represents a unified model of functional coupling between *Oryza sativa* genes.

The next sections describe those aspects of network reconstruction that differ from (1).

**References and benchmark sets used for RiceNet construction and evaluation**

The Gene Ontology (GO) biological process (BP) annotation from TIGR Rice Genome Annotation Release 5 (4) served as the major reference set for training and benchmarking the network. To minimize functional bias in the training set, we excluded annotations made directly to the following terms: 1) 6 dominant terms (these 6 terms out of more than 1,000 BP terms account for >60% of total training gene pairs and were thus removed), "defense response" (GO:0006952), "protein amino acid phosphorylation" (GO:0006468), "regulation of transcription, DNA-dependent" (GO:0006355), "regulation of transcription" (GO:0045449), "proteolysis" (GO:0006508), "transmembrane receptor protein tyrosine kinase signaling pathway" (GO:0007169); 2) 3 additional terms that include highly diverse biological processes, "hypersensitive response" (GO:0009626), "signal transduction" (GO:0007165), and "transport" (GO:0006810), and 3) 4 direct children of the BP root term, "metabolic process" (GO:0008152), "growth" (GO:0040007), "photosynthetic water oxidation" (GO:0009781), and "reproduction" (GO:0000003). The resulting dataset of reference gene annotations for training contained 662,936 pairs covering 8,716 *Oryza sativa* genes (~21% of 41,203 non-TE-related protein-coding loci). For validating RiceNet using independent annotation, we employed the Kyoto-based KEGG pathway database (5). Links generated between genes sharing KEGG annotation terms overlap minimally (2%) with the GO-BP training set.

**Inferring functional linkages from mRNA expression data**

One major source of functional associations was mRNA co-expression data. We inferred functional associations from mRNA co-expression patterns largely as previously described (6), with the following modifications: DNA microarray data for *Oryza sativa* genes was downloaded from the Gene Expression Omnibus (GEO) (7) on August 2009. Data were first filtered to remove uninformative sets by testing for a significant correlation between the Pearson correlation coefficient (PCC) between pairs of genes' expression vectors and the genes' sharing of GO biological processes (measured as LLS scores). Out of 23 data sets with at least 10 experiments each (comprising a total of 718 microarray experiments), we found only 11 data sets comprising 274 experiments (**Supplementary Table 1A**) showed a significant correlation between co-expression and the tendency for genes to belong the same processes. Gene linkages derived from each of these 11 DNA microarray experiment sets then were integrated by the weighted-sum

method as described above. The remaining 12 data sets (comprising 444 microarray experiments; **Supplementary Table 1B**) did not show such a correlation and were thus omitted from further analysis.

**Inferring functional linkages from the genomic context of orthologous proteins**

We inferred functional associations between rice genes from comparative genomics analyses as follows. Both phylogenetic profiling (8-10) and gene neighbors (11-13) among prokaryotic orthologs of *O. sativa* genes performed reasonably well at identified rice gene functional associations, calculating links as in (1). We observed the best performance (judged by recall-precision analysis) for calculations based on a set of 184 unique bacterial and archeael genera previously identified in (1). Representative genomes for each unique genus were chosen according to which species within the genus exhibited the largest number of proteins hit by BLASTP in the *O. sativa* proteome. We assigned log likelihood scores to each *O. sativa* gene pair based upon a regression model relating the LLS to the mutual information between the phylogenetic profiles or to the log of the probability of observing gene neighbors by chance, calculated as in (3).

**Inferring functional linkages from associalogs**

RiceNet includes many linkages transferred from other organisms' gene networks where the linked genes are both conserved. Such conserved functional associations (termed associalogs (3)) provide a valuable source of data in rice, where primary rice-specific datasets are still quite limiting. Associalogs were scored as for any other *O. sativa* dataset (e.g., assigning LLS scores to the transferred linkages using the *O. sativa* annotation benchmarks), with the additional steps of calculating orthologs and weighting linkages by the confidence in the orthology assignments. We identified orthologs using INPARANOID (14), and learned rice gene linkages based on linkages from version 3 of YeastNet (6), version 2 of WormNet (3), and version 1 of HumanNet (15) and AraNet (1). For calculating associalogs, we used the scheme of (1), in which each type of evidence in a second organism (mRNA co-expression, yeast two-hybrid interactions, and etc.) was treated as an individual data set. A total of 21 associalog linkage sets were generated: 3 from *Arabidopsis*, 5 from worm, 5 from human, and 8 from yeast. We weighted orthology-based functional inferences by the INPARANOID confidence scores using the scheme of (1). Each such associalog dataset was then scored as for other *O. sativa* datasets, using a regression model between the weighted scores and the LLS for sharing *O. sativa* functional annotation.

**Construction of MaizeNet by orthology-based links from AraNet and RiceNet**

In order to test the power of RiceNet and AraNet for predicting maize Gene Ontology biological process terms, we constructed networks of maize genes (MaizeNet) by transferring links from AraNet and RiceNet by orthology, using the same methods as described above for weighting functional links from associalogs. We defined a RiceNet-derived MaizeNet (OS-ZM) with 788,469 links, each at least 1.5 times more likely than random expectation. We also defined an AraNet-derived MaizeNet (AT-ZM) with the same number of links as in OS-ZM. The two networks, OS-ZM and AT-ZM, contain 17,662 and 15,167 genes, respectively.

**ROC analysis of gene function**

The ability of each gene network model to predict gene function was tested using Receiver Operating Characteristic (ROC) plots and leave-one-out cross-validation as follows. For each set of query genes associated with the same known function, we rank-ordered all rice genes (including the query genes) by their connectivity to the query genes using the method of Gaussian field label propagation (16) as implemented in (17). The quality of predictions was then assessed by Receiver Operating Characteristic (ROC) analysis, plotting true positive versus true negative prediction rates. In essence, this measures the tendency of genes with the same function to cluster in the network, and whether this tendency is useful for ranking genes most likely to share that function. For cases in which genes annotated to have the same function cluster tightly in the network, we expect a higher retrieval rate for query genes (positives) than for non-query genes (negatives) in a ROC plot, resulting in a ROC curve above the diagonal. However, if the genes known to be involved in the same function are not clustered in the network, the retrieval rate of positive and negative genes will be similar, resulting in a diagonal ROC curve, indicating random prediction. Each such ROC analysis can be summarized by the area under the ROC curve (AUC), which ranges from near 0.5 (i.e., the area under the diagonal, indicating random performance) to 1 (genes with this function are completely connected in the network).

**Analysis of cell-type specific expression specificity**

In order to assess the similarity of cell-type expression profiles for linked genes, we first generated a set of strongly-expressed genes in each of 40 *O. sativa* cell types, using for this purpose the data for shoots, roots and germinating seeds at several developmental stages collected by Jiao *et al*. (18). We selected those transcripts with DNA microarray-based integrated signal intensities ≥ 2,000, resulting in roughly 1,000-2,000 strongly

expressed genes per cell type. (Note that limiting the analysis to well expressed genes merely serves to lower the random expectation of co-expression across cell types, hence giving a more sensitive test.) We measured the tendency for linked genes to be co-expressed across cell types as an odds ratio, calculated as in (1).

**Plasmid construction for yeast two-hybrid (Y2H) experiments and generation of transgenic plants**

The amplified full-length cDNAs for 16 RiceNet candidates from a rice cDNA or KOME (Knowledge-based Oryza Molecular biological Encyclopedia) cDNA (19) using specific primers (described in **Supplementary Table 8)** were cloned into pENTR/D TOPO (Invitrogen) according to the instructions provided by the manufacturer and inserts confirmed by sequencing. Positive clones were verified by sequencing and then moved by Gateway LR Clonase (Invitrogen) into the yeast two hybrid vector pNlexA carrying the BD domain (Clontech) or pB42AD containing the AD domain (Clontech) for yeast two hybrid assays. In addition, five positive clones were moved into the Ubi-C1300 binary vector (20) in order to generate overexpression transgenic lines. Partial fragments (350-450bp) of four RiceNet candidates were amplified from rice cDNA or KOME cDNA using primers (described in **Supplementary Table 8)**, cloned and recombined into the pANDA binary vector for generation of RNAi silencing transgenic lines using the same procedure.

**Yeast two-hybrid (Y2H) analysis**

Detailed experimental procedures are described in Seo *et al.* (21). Briefly, the purified plasmid DNAs from BD vector-bait genes and AD vector-prey genes were transformed into yeast strain pEGY48/p8op-LacZ (Clontech). The strength of each protein-protein interaction was determined by color development of a yeast colony two days after streaking on medium containing X-Gal. For a negative interaction control, we employed an empty vector as one of the interaction partners. For a positive control, we used XA21-XB3, a known strong interactor described in (21). An interaction was considered significant and reproducible if we observed similar results from 2 to 3 independent assays. For an estimate of random interactions, we tested a protein (SP11) that was not predicted to be connected to the 24 XA21 interactome proteins in RiceNet and observed that SP11 (S-locus Cys-rich protein, located in the pollen coat and which determines pollen S-haplotype specificity) did not interact with the 24 tested XA21 interactome proteins. Thus, we expect a very low background probability of interaction of rice proteins to the tested 24 XA21 interactome proteins. The results of yeast two hybrid

analysis are summarized in **Supplementary Table 5.**

**Generation of transgenic plants**

In order to validate RiceNet-guided predictions, we used a previously generated transgenic rice line carrying the *Xa21* gene under its native promoter in the Kitaake genetic background. Rice transformation was carried out as described previously (22). All generated transgenic lines were tested and confirmed by PCR with gene-specific and/or construct-specific primers. We generated at least 3-7 independent transformed plants for each line except for the Xa21-LOC_Os02g21510 RNAi line.

**Expression quantification and *Xoo* inoculation methods**

Altered expression of the transgenes in the transgenic lines was validated by semi-quantitative reverse transcription PCR (**Supplementary Figure 3a, 3c, and 3e**). For *Xoo* inoculation, a bacterial suspension ($OD_{600}$ of 0.5) of *Xoo* strain PXO99, which carries Ax21 activity was used. Five week-old transgenic plants were inoculated with *Xoo* via the scissors-dip method (23).

**Genetic analyses of ROX1, ROX2 and ROX3 transgenic plants**

***Rox1:*** T1 progeny containing the LOC_Os01g70580 RNAi construct displayed enhanced susceptibility to *Xoo*(lesion lengths of *3-7 cm* +/- 1.3 cm) as compared with the Kitaake-XA21 control (*2 cm* +/- 0.8 cm) (**Supplementary Figure 3b**). Leaf lesion lengths correspond to bacterial growth *in planta* (22, 23). We next self-pollinated those T1 lines (2-2, 2-7, 2-10, and 2-11) that displayed highly enhanced susceptibility to *Xoo* as compared with the Kitaake-XA21 control. The resulting T2 progeny lines segregated for the enhanced susceptibility phenotypes. The resulting T2 transgenic plants carrying the LOC_Os01g70580 RNAi construct (2-2, 2-7, 2-10, and 2-11) displayed enhanced susceptibility to *Xoo* (lesion lengths of *7-11 cm* +/- 1.5 cm), whereas the T2 progeny 2-7-4 and 2-7-10 that do not contain the LOC_Os01g70580 RNAi construct displayed no phenotypic alteration (lesion lengths of *2-3* cm +/- 0.5 cm) (**Figure 3a**). The knock-down of gene expression was correlated with the enhanced susceptibility phenotype in LOC_Os01g70580 RNAi 2-10 and 2-11 lines, with all RNAi lines with decreased target gene expression showing enhanced susceptibility (**Supplementary Figure 3a-b**). These results indicate that LOC_Os01g70580 (now designated *Rox1*, Regulator of XA21-mediated immunity) is a positive regulator of XA21-mediated immunity.

***Rox2:*** From the three independent XA21-LOC_Os02g21510 RNAi lines generated, T1 progeny from two of the lines (T0 lines 3 and 4) displayed decreased accumulation

of LOC_Os02g21510 mRNA. The reduced mRNA accumulation correlated with susceptibility to *Xoo* (lesion lengths of *4-5 cm* +/- 0.5 cm) (**Supplementary Figure 3c-d**) in these lines. Control Kitaake-XA21 displayed shorter lesions (lesion lengths of *2 cm* +/- 0.6 cm) (**Supplementary Figure 3d**). T1 progeny from two lines (3 and 4) as well as T2 progeny from line 3-1 expressing the LOC_Os02g21510 RNAi construct displayed enhanced susceptibility to *Xoo* (lesion lengths of 4-7 cm +/- 1.1 cm) (**Figure 3b**). These results indicate that LOC_Os02g21510, designated *ROX2*, is also a positive regulator of XA21-mediated immunity.

***Rox3:*** We next investigated the resistance response of three independent XA21-LOC_Os06g12530 overexpression lines. Two T1 progeny (from T0 lines 1 and 3) were selected based on an observed increased accumulation of LOC_Os06g12530 mRNA in these lines. These lines displayed enhanced susceptibility to *Xoo* (lesion lengths of *4-5 cm* +/- 0.5 cm) (**Supplementary Figure 3e-f**). T1 progeny from two T0 lines (1 and 3) and T2 progeny from one T1 line (1-15) carrying the LOC_Os06g12530 overexpression construct exhibited enhanced susceptibility (lesion lengths of4-8 cm +/- 1.3 cm) compared with XA21 plants *(*lesion lengths of 1.8 cm +/- 0.7 cm) (**Figure 3c**). These results indicate that LOC_Os06g12530, designated *ROX3*, serves as a negative regulator of XA21-mediated immunity.

**Supplementary Table 1A**. Rice mRNA expression datasets incorporated into RiceNet. Accession numbers reference the GEO database.

| Accession | Set title | # array | # links |
|---|---|---|---|
| GSE4409 | Dissecting the rice genes responsible for long time changes of nitrogen supply forms and nitrogen starvation | 12 | 40,334 |
| GSE4438 | Expression data from rice under salinity stress (24) | 24 | 146,983 |
| GSE6893 | Expression data for reproductive development in rice (25, 26) | 45 | 154,076 |
| GSE6901 | Expression data for stress treatment in rice seedlings (25, 26) | 12 | 19,764 |
| GSE7071 | Reactive oxygen species trigger a regulatory module involved in the early responses of rice seedlings to cold stress (27) | 20 | 4,048 |
| GSE7531 | Differentially expressed genes from different developmental stages in rice | 26 | 4,095 |
| GSE7532 | Differentially expressed genes under different stress conditions in rice | 26 | 30,196 |
| GSE7951 | Genome-wide gene expression profiling of rice stigma (28) | 13 | 151,828 |
| GSE10373 | Rice cultivars undergoing a susceptible and resistant interaction with the parasitic plant Strigahermonthica (29) | 24 | 97,329 |
| GSE11157 | Microarray analysis of rice plants fumigated with ozone (30) | 12 | 51,227 |
| GSE16793 | Comparative transcriptional profiling of rice undergoing infection by *X. oryzaepv. Oryzae* or by *X. oryzaepv. oryzicola* | 60 | 145,986 |

**Supplementary Table 1B**. Rice mRNA expression datasets lacking correlation between mRNA co-expression and LLS scores.

| Accession | Set title | # array |
|---|---|---|
| GSE2415 | Rice seedling stress treatment | 12 |
| GSE5853 | Transcriptomic adaptations in rice suspension cells under sucrose starvation (31) | 12 |
| GSE6719 | Cytokinin responsive genes in rice (32) | 24 |
| GSE7530 | Gene expression profiles of various tissues under different physiological conditions | 39 |
| GSE7766 | Moderate long-term drought stress in rice | 14 |
| GSE8518 | Rice Gene Expression During Biotrophic Invasion by *Magnaportheoryzae* (33) | 12 |
| GSE9450 | Rice plants infected by Rice blast fungus (34) | 36 |
| GSE9653 | NSF 20K Oligo Arrays to Dissect Rice defense response pathways | 114 |
| GSE10857 | Gene expression of rice root tips before, at and buckled by a hard layer in two rice varieties | 12 |
| GSE11896 | Global transcriptome profiles in the nodes of rice during spatial and temporal development | 134 |
| GSE14403 | Root-specific transcriptional profiling of contrasting rice genotypes in response to salinity stress | 23 |
| GSE15046 | Transcriptome analysis of gibberellin-signaling mutants in rice (35) | 12 |

**Supplementary Table 2**. A list of twenty-four data types incorporated into RiceNet

| Evidence code | Evidence description | # unique genes | # unique gene pairs |
|---|---|---|---|
| AT-CX | Co-expression among *A. thaliana* genes | 4,488 | 81,671 |
| AT-DC | Co-occurrence of domains among *A. thaliana* proteins | 3,930 | 10,909 |
| AT-LC | Literature curated *A. thaliana* protein physical interactions | 614 | 786 |
| CE-CC | Co-citation of *C. elegans* orthologs in Medline abstracts | 920 | 9,093 |
| CE-CX | mRNA co-expression of *C. elegans* orthologs | 2,434 | 50,824 |
| CE-GT | Genetic interactions between *C. elegans* orthologs | 574 | 3,000 |
| CE-LC | Literature curated *C. elegans* protein physical interactions | 970 | 1,991 |
| CE-YH | High-throughput yeast 2-hybrid interactions among *C. elegans* orthologs | 1,050 | 2,122 |
| HS-CX | mRNA co-expression between human orthologs | 3,039 | 39,561 |
| HS-DC | Co-occurrence of domains among human proteins | 4,256 | 29,288 |
| HS-LC | Literature curated human protein physical interactions | 3,828 | 85,926 |
| HS-MS | Human protein complexes from affinity purification/mass spectrometry | 1,007 | 2,909 |
| HS-YH | High-throughput yeast 2-hybrid interactions among human orthologs | 676 | 1,239 |
| OS-CX | Co-expression among *O. sativa* genes | 9,604 | 39,415 |
| OS-GN | Gene neighborhoods of prokaryotic orthologs of *O. sativa* genes | 5,352 | 106,122 |
| OS-PG | Co-inheritance of prokaryotic orthologs of *O. sativa* genes | 1,810 | 42,107 |
| SC-CC | Co-citation of yeast orthologs in Medline abstracts | 3,735 | 38,642 |
| SC-CX | mRNA co-expression among yeast orthologs | 2,804 | 98,708 |
| SC-DC | Co-occurrence of domains among yeast proteins | 2,849 | 15,767 |
| SC-GT | Genetic interactions between yeast orthologs | 3,042 | 18,763 |
| SC-LC | Literature curated yeast protein physical interactions | 2,529 | 12,113 |
| SC-MS | Yeast protein complexes from affinity purification/mass spectrometry | 3,140 | 100,428 |
| SC-TS | Yeast protein interactions inferred from tertiary structures of complexes | 1,230 | 6,221 |
| SC-YH | High-throughput yeast 2-hybrid interactions among yeast orthologs | 2,213 | 6,556 |

**Supplementary Table 3**. XA21-interactome proteins. To identify candidate genes governing the rice defense response, we queried RiceNet with 15 member proteins of the XA21 interactome with clear phenotypes (bold italics; see **Figure 2**). To prioritize genes for detailed analysis, we tested the interaction of RiceNet-predicted proteins for interaction with XA21 and 23 members of the XA21 interactome (shaded rows).

| **Locus ID** | **Name** | **Locus ID** | **Name** |
|---|---|---|---|
| LOC_Os01g03820 | XB11 | LOC_Os05g28300 | XB12IP5 |
| ***LOC_Os01g12900*** | ***OsRac1*** | ***LOC_Os05g45420*** | ***SnRK1A*** |
| LOC_Os01g14810 | LOC_Os01g14810 | LOC_Os05g49700 | XB12IP1 |
| LOC_Os01g25820 | OsRBOHB | LOC_Os06g06090 | OsMPK1 |
| LOC_Os01g32660 | OsMKK6 | LOC_Os06g17280 | XB2 |
| LOC_Os01g47530 | OsMPK8 | LOC_Os06g44010 | OsWRKY28 |
| LOC_Os01g49290 | RACK1 | LOC_Os06g46770 | XB2IP1 |
| ***LOC_Os01g56470*** | ***XB24*** | ***LOC_Os06g49430*** | ***OsMPK12*** |
| LOC_Os02g05490 | XB2IP4 | LOC_Os07g07540 | XB22 |
| ***LOC_Os02g08440*** | ***OsWRKY71*** | LOC_Os08g05560 | XB11IP2 |
| ***LOC_Os02g33180*** | ***RAR1*** | ***LOC_Os08g34740*** | ***SGT1*** |
| ***LOC_Os02g54160*** | ***OsEREBP-1*** | LOC_Os09g04810 | XB2IP3 |
| LOC_Os02g54600 | OsMKK4 | ***LOC_Os09g25060*** | ***OsWRKY76*** |
| LOC_Os02g57200 | LOC_Os02g57200 | ***LOC_Os09g25070*** | ***OsWRKY62*** |
| LOC_Os03g08550 | PBZ1IP-1 | LOC_Os09g30412 | HSP90 |
| ***LOC_Os03g12470*** | ***WAK25*** | LOC_Os10g28610 | KIP1 |
| ***LOC_Os03g17700*** | ***OsMPK5*** | ***LOC_Os10g38950*** | ***OsMPK6*** |
| LOC_Os03g45280 | PBZ1IP-2 | LOC_Os10g42700 | XB12 |
| ***LOC_Os03g60650*** | ***XB15*** | LOC_Os11g01550 | XB22IP1 |
| LOC_Os04g39090 | XB12IP2 | LOC_Os11g36070 | XB11IP1 |
| LOC_Os04g55480 | XB2IP2 | LOC_Os12g36180 | XB21 |
| ***LOC_Os05g02130*** | ***XB3*** | LOC_Os12g36880 | PBZ1 |
| LOC_Os05g09020 | OsWRKY67 | No Locus ID | XA21 |

**Supplementary Table 4**. Identification of fourteen genes predicted to be involved in XA21-mediated immunity. Five of these (shaded rows) were further prioritized based on their direct interactions with XA21 (LOC_Os01g70580, LOC_Os01g70790, LOC_Os02g21510, and LOC_Os03g20460) or because they contain a sequence motif typical of genes governing the animal inflammatory response (LOC_Os06g12530). Expression of each of the five genes was altered through transgenic analysis, and the resulting mutant lines were tested for altered immune response. Three out of five genes (bold italics) displayed altered immune responses, corresponding to a 60% success rate. Data types supporting association of each candidate gene to the query genes are indicated by evidence codes as in **Figure 1a**, accompanied by the weighted contribution of that line of evidence to the total support for that candidate gene.

| Locus ID | Score | #RiceNet links to the set of 15 query genes | Supporting data type: Probability of contribution |
|---|---|---|---|
| *LOC_Os01g47770* | 2.2 | 1 | HS-DC:1.00 |
| ***LOC_Os01g70580*** | ***1.72*** | ***3*** | ***OS-GN:1.00*** |
| *LOC_Os01g70790* | 2.45 | 1 | AT-CX:1.00 |
| *LOC_Os02g05480* | 2.09 | 3 | SC-CC:0.63 SC-TS:0.37 |
| *LOC_Os02g15810* | 4.15 | 3 | HS-LC:0.71 AT-CX:0.29 |
| ***LOC_Os02g21510*** | ***1.64*** | ***3*** | ***OS-GN:1.00*** |
| *LOC_Os03g07300* | 2.67 | 4 | OS-GN:1.00 |
| *LOC_Os03g20460* | 1.47 | 2 | OS-GN:1.00 |
| *LOC_Os03g26460* | 2.2 | 1 | HS-DC:1.00 |
| *LOC_Os03g53720* | 2.73 | 3 | SC-TS:0.74 HS-CX:0.26 |
| *LOC_Os04g35700* | 2.32 | 2 | SC-CC:1.00 |
| *LOC_Os05g27730* | 3.13 | 1 | OS-CX:0.58 AT-CX:0.42 |
| ***LOC_Os06g12530*** | ***2.2*** | ***1*** | ***HS-DC:1.00*** |
| *LOC_Os07g48290* | 2.74 | 3 | SC-TS:0.47 HS-LC:0.36 SC-MS:0.16 |

**Supplementary Table 5**. Results of yeast two hybrid tests between XA21 interactome proteins and candidate XA21 regulators predicted by RiceNet.

| XA21 interactome protein | Colony color intensity* | XA21 regulator candidate | XA21 interactome protein | Colony color intensity* | XA21 regulator candidate |
|---|---|---|---|---|---|
| XB12 | +++ | LOC_Os02g15810 | XA21 | ++ | LOC_Os03g20460 |
| XB12 | +++ | LOC_Os02g21510 | XB22 | ++ | LOC_Os05g27730 |
| XB12 | +++ | LOC_Os03g07300 | XB12 | ++ | LOC_Os06g12530 |
| OsWRKY62 | +++ | LOC_Os03g20460 | XA21 | + | LOC_Os01g70580 |
| XB12 | +++ | LOC_Os03g20460 | OsWRKY71 | + | LOC_Os01g70580 |
| XB15 | +++ | LOC_Os03g20460 | SnRK1A | + | LOC_Os01g70580 |
| XB24 | +++ | LOC_Os03g20460 | XB2 | + | LOC_Os01g70790 |
| XB22IP-1 | +++ | LOC_Os03g20460 | HSP90 | + | LOC_Os01g70790 |
| HSP90 | +++ | LOC_Os03g26460 | XB11IP-2 | + | LOC_Os01g70790 |
| XB12 | +++ | LOC_Os03g53720 | XB12IP-1 | + | LOC_Os01g70790 |
| SnRK1A | +++ | LOC_Os04g35700 | OsEREBP1 | + | LOC_Os01g70790 |
| HSP90 | +++ | LOC_Os01g47770 | SnRK1A | + | LOC_Os02g15810 |
| OsWRKY62 | +++ | LOC_Os01g70790 | XB12IP-1 | + | LOC_Os02g15810 |
| XB12 | +++ | LOC_Os01g70790 | OsEREBP1 | + | LOC_Os03g20460 |
| XB15 | +++ | LOC_Os01g70790 | XB12 | + | LOC_Os03g26460 |
| XB21 | +++ | LOC_Os01g70790 | XB24 | + | LOC_Os03g26460 |
| SnRK1A | +++ | LOC_Os01g70790 | SnRK1A | + | LOC_Os03g26460 |
| XB22IP-1 | +++ | LOC_Os01g70790 | XB22IP-1 | + | LOC_Os03g26460 |
| XB3 | ++ | LOC_Os01g70790 | SnRK1A | + | LOC_Os03g53720 |
| XB24 | ++ | LOC_Os01g70790 | OsWRKY71 | + | LOC_Os04g35700 |
| OsWRKY71 | ++ | LOC_Os01g70790 | XB12IP-2 | + | LOC_Os04g35700 |
| XB12IP-2 | ++ | LOC_Os01g70790 | SnRK1A | + | LOC_Os05g27730 |
| XA21 | ++ | LOC_Os01g70790 | XB12IP-1 | + | LOC_Os06g12530 |
| OsWRKY62 | ++ | LOC_Os02g15810 | XB12 | + | LOC_Os07g48290 |
| XB24 | ++ | LOC_Os02g15810 | OsWRKY71 | + | LOC_Os01g47770 |
| XA21 | ++ | LOC_Os02g21510 | XB12IP-1 | + | LOC_Os01g47770 |
| XB3 | ++ | LOC_Os03g20460 | XB22IP-1 | + | LOC_Os01g47770 |
| SnRK1A | ++ | LOC_Os03g20460 | OsEREBP1 | + | LOC_Os01g47770 |
| XB12IP-1 | ++ | LOC_Os03g20460 | | | |

* Strength of protein-protein interaction is indicated by intensity of colony color development; +, weak; ++, medium; +++, strong.

**Supplementary Table 6**. Lesion length of transgenic plants (Kitaake-XA21 genetic background) inoculated with the bacterial pathogen *Xanthomonas oryzae pv. oryzae (Xoo)*, No obvious phenotypic changes were observed in transgenic lines that over-express LOC_Os01g70580 (ROX1), LOC_Os02g21510 (ROX2), LOC_Os01g70790, LOC_Os03g20460, or those that are silenced for LOC_Os01g70790 and LOC_Os03g20460.

| Transgenic lines (Kitakke-XA21 background) | Disease lesion length (Average ± SD) | Genotype |
|---|---|---|
| LOC_Os01g70580 overexpression line-1 | 1.9 ± 1.1 | (+) |
| LOC_Os01g70580 overexpression line-2 | 2.0 ± 1.4 | (+) |
| LOC_Os01g70580 overexpression line-3 | 2.5 ± 0.7 | (-) |
| LOC_Os01g70580 overexpression line-4 | 2.9 ± 1.5 | (+) |
| LOC_Os01g70580 overexpression line-5 | 3.4 ± 1.6 | (+) |
| LOC_Os01g70580 overexpression line-6 | 1.3 ± 0.4 | (+) |
| LOC_Os01g70580 overexpression line-7 | 3.3 ± 1.7 | (+) |
| LOC_Os01g70580 overexpression line-8 | 1.4 ± 0.6 | (+) |
| LOC_Os01g70580 overexpression line-9 | 0.8 ± 0.4 | (+) |
| Kitakke (control) | 13.5 ± 2.1 | |
| Kitakke-XA21 (control) | 2.7 ± 1.1 | |
| LOC_Os02g21510 overexpression line-1 | 2.4 ± 1.0 | (-) |
| LOC_Os02g21510 overexpression line-2 | 2.4 ± 0.6 | (+) |
| LOC_Os02g21510 overexpression line-3 | 1.8 ± 0.4 | (+) |
| LOC_Os02g21510 overexpression line-4 | 2.0 ± 0.7 | (+) |
| LOC_Os02g21510 overexpression line-5 | 1.3 ± 0.4 | (+) |
| LOC_Os02g21510 overexpression line-6 | 3.0 ± 0.7 | (+) |
| LOC_Os02g21510 overexpression line-7 | 1.5 ± 0.6 | (+) |
| Kitakke (control) | 12.5 ± 1.1 | |
| Kitakke-XA21 (control) | 1.7 ± 1.0 | |
| LOC_Os01g70790 overexpression line-1 | 0.5 ± 0.7 | (+) |
| LOC_Os01g70790 overexpression line-2 | 0.5 ± 0.4 | (+) |
| LOC_Os01g70790 overexpression line-3 | 1.5 ± 0.6 | (+) |
| LOC_Os01g70790 overexpression line-5 | 1.8 ± 0.4 | (+) |
| LOC_Os01g70790 overexpression line-6 | 3.5 ± 0.7 | (+) |
| LOC_Os01g70790 overexpression line-7 | 2.5 ± 1.1 | (-) |
| LOC_Os01g70790 RNAi line-1 | 0.8 ± 0.4 | (+) |
| LOC_Os01g70790 RNAi line-2 | 1.1 ± 0.2 | (+) |

| | | |
|---|---|---|
| LOC_Os01g70790 RNAi line-3 | 2.3 ± 1.0 | (+) |
| Kitakke (control) | 15.0 ± 2.7 | |
| Kitakke-XA21 (control) | 2.2 ± 1.0 | |
| LOC_Os03g20460 overexpression line-1 | 1.0 ± 0.2 | (+) |
| LOC_Os03g20460 overexpression line-2 | 1.0 ± 0.0 | (+) |
| LOC_Os03g20460 overexpression line-3 | 1.9 ± 0.5 | (+) |
| LOC_Os03g20460 RNAi line-1 | 1.5 ± 0.3 | (+) |
| LOC_Os03g20460 RNAi line-2 | 3.0 ± 1.1 | (+) |
| LOC_Os03g20460 RNAi line-3 | 1.0 ± 0.4 | (+) |
| LOC_Os03g20460 RNAi line-4 | 2.5 ± 0.0 | (+) |
| LOC_Os03g20460 RNAi line-5 | 2.8 ± 1.6 | (+) |
| LOC_Os03g20460 RNAi line-6 | 3.0 ± 1.0 | (+) |
| LOC_Os03g20460 RNAi line-7 | 0.5 ± 0.0 | (+) |
| Kitakke (control) | 14.0 ± 2.5 | |
| Kitakke-XA21 (control) | 1.5 ± 1.3 | |

**Supplementary Table 7.** Predictive strength of the AraNet-derived MaizeNet (AT-ZM) and the RiceNet-derived MaizeNet (OS-ZM) for maize Gene Ontology biological processes. For the 30 GO biological processes predictable (AUC > 0.5) by either network, OS-ZM performed better than AT-ZM for 20 processes (shaded rows).

| Maize GO-BP terms | AUC by AT-ZM | AUC by OS-ZM | GO term size |
|---|---|---|---|
| response to oxidative stress | 0.78 | 1.00 | 3 |
| response to osmotic stress | 1.00 | 1.00 | 4 |
| protein homotetramerization | 0.78 | 0.84 | 9 |
| response to herbicide | 0.77 | 0.84 | 9 |
| response to cold | 0.69 | 0.82 | 6 |
| response to heat | 0.82 | 0.81 | 7 |
| response to ethylene stimulus | 0.78 | 0.77 | 3 |
| photosynthesis | 1.00 | 0.77 | 3 |
| sequestering of actin monomers | 0.55 | 0.77 | 3 |
| cytokinin mediated signaling | 0.67 | 0.77 | 6 |
| response to abscisic acid stimulus | 0.68 | 0.73 | 12 |
| response to salt stress | 0.66 | 0.72 | 4 |
| cold acclimation | 0.68 | 0.72 | 5 |
| response to auxin stimulus | 0.36 | 0.70 | 6 |
| response to light stimulus | 0.72 | 0.70 | 9 |
| circadian rhythm | 0.54 | 0.68 | 4 |
| protein folding | 0.69 | 0.67 | 5 |
| response to reactive oxygen species | 0.47 | 0.64 | 11 |
| protein heterotetramerization | 0.70 | 0.63 | 9 |
| response to copper ion | 0.55 | 0.61 | 3 |
| response to hydrogen peroxide | 0.46 | 0.61 | 5 |
| response to wounding | 0.58 | 0.59 | 6 |
| nuclear migration | 0.73 | 0.58 | 5 |
| two-component signal transduction system (phosphorelay) | 0.65 | 0.57 | 9 |
| response to fungus | 0.50 | 0.56 | 9 |
| response to salicylic acid stimulus | 0.45 | 0.56 | 7 |
| response to xenobiotic stimulus | 0.43 | 0.55 | 15 |
| response to gibberellic acid stimulus | 0.33 | 0.55 | 3 |
| response to cytokinin stimulus | 0.52 | 0.52 | 25 |
| Defense response to fungus | 0.47 | 0.46 | 9 |
| response to jasmonic acid stimulus | 0.58 | 0.45 | 6 |
| Response to nitrate | 0.33 | 0.30 | 4 |

**Supplementary Table 8.** Sequences of forward (F) and reverse (R) primers used for cloning cDNAs

| Our name | TIGR Locus ID | Primer sequence |
|---|---|---|
| XB2a | LOC_Os06g17280 | F 5 CCCGGCCGCTTCCCTTCC 3<br>R 5 TCACTCCATCTTCCTGCCAGA 3 |
| XB3 | LOC_Os05g02130 | F 5 ATGGGTCACGGTGTCAGC 3<br>R 5 TCATAGATCGTGCTCAGGCTT 3 |
| XB10 | LOC_Os09g25070 | F 5 ATGGACGACGACGGCGAC 3<br>R 5 CTACAAATGAACAGGAATGTG 3 |
| XB11 | LOC_Os01g03820 | F 5 ATGGCGAAGAAGAAGCTGAAG 3<br>R 5 TCACCACCAAACCTTAGTTCG 3 |
| XB12 | LOC_Os10g42700 | F 5 ATGTTGGCCTCCACTTGCTTC 3<br>R 5 TCATATGTTGGTGGGTGGTGC 3 |
| XB15 | LOC_Os03g60650 | F 5 ATGGGCAACTCCCTCGCC 3<br>R 5 TTACACGCAGGATCTCCAAAT 3 |
| XB21 | LOC_Os12g36180 | F 5 ATGGACGACTTCCAGGGCCTC 3<br>R 5 TTAGAAGAGTTCCTCTGAGTTGAA 3 |
| XB22 | LOC_Os07g07540 | F 5 ATGGCGCTCGCCCACCAG 3<br>R 5 TCATTTCTTGTTGAATCCAAACAA 3 |
| XB24 | LOC_Os01g56470 | F 5 ATGGGTTGGCGTTGGCAC 3<br>R 5 TTACACATCTGTAATCTTGCTGCT 3 |
| HSP90 | LOC_Os09g30412 | F 5 ATGGCGTCGGAGACCGAG 3<br>R 5 TTAGTCGACCTCCTCCATCTT 3 |
| RAR1 | LOC_Os02g33180 | F 5 ATGTCGACGGAGGCGGAG 3<br>R 5 TCATGCGGCATCAGCATTGTG 3 |
| SGT1 | LOC_Os01g43540 | F 5ATGGCGACCTCCGCCTCC 3<br>R 5 TCACGGCTTTTTACCATCAGG 3 |
| XB2IP-2 | LOC_Os04g55480 | F 5 ATGTTCAGCCTTCTGATCCAG 3<br>R 5 CTACATTGTCAGTGCTGCCTC 3 |
| XB10IP-1 | LOC_Os02g08440 | F 5 ATGGATCCGTGGATTAGCACC 3<br>R 5 TCAATCCTTGGTCGGCGAGAG 3 |
| SNF1 | LOC_Os05g45420 | F 5 ATGGAGGGAGCTGGCAGAGAT 3<br>R 5 TTAAAGGACTCTCAGCTGAGT 3 |
| XB11IP-1 | LOC_Os11g36070 | F 5 ATGGTGGCGGCCATGGAG 3<br>R 5 TCATCCTGTGGTGCAAGGGGT 3 |
| XB11IP-2 | LOC_Os08g05560 | F 5 ATGACCGGCGCCGCCGCA 3<br>R 5 TCATCTCTTCCCAAAGAAATG 3 |
| LIP5 | LOC_Os05g45420 | F 5 ATGGCCGGCATCATCCA 3<br>R 5 TCAGTCGCTGTCGCTGCTGCT 3 |
| XB12IP-1 | LOC_Os11g36070 | F 5 ATGGACGCCAGCCTCCGC 3<br>R 5 CTAGAAGTTAAGCATCTCCCA 3 |
| XB12IP-2 | LOC_Os08g05560 | F 5 ATGGCCTCCTCGGCGTC 3<br>R 5 TCATCTACGCCGTTTTGAATAAGG 3 |
| XB22IP-1 | LOC_Os11g01550 | F 5ATGGCGGGAAGCGGGAGC 3<br>R 5 TTATGTCCACATGGACTCTTT 3 |
| XB22IP-2 | LOC_Os02g54160 | F 5ATGTGCGGCGGCGCCATC 3<br>R 5 TCAATAGAAATCGCTAACGGG 3 |
| OsMPK12 | LOC_Os06g49430 | F 5 ATGGGGGGAGGGGGCACG 3<br>R 5 CTAGGAGTGCATCCTGGAGAC 3 |
| $N2^{ox}$ | LOC_Os01g70580 | F 5 ATGCCGCTGCCGACGATGACC 3<br>R 5 TCAGTTCCGGAGTGATATTGTCCA 3 |

| | | |
|---|---|---|
| N3 [ox] | LOC_Os01g70790 | F 5 ATGGCGCAGAGGACGCTGGA 3 |
| | | R 5 CTAGAAATCGAAGCCACCGCC 3 |
| N4K | LOC_Os02g15810 | F 5 ATGCAGAGGACCGCCAAGGAG 3 |
| | | R 5 CTAGGACGAGGCCGAGGAGG 3 |
| N6 [ox] | LOC_Os02g21510 | F 5 ATGGCCGACGCTAGGTCCGC 3 |
| | | R 5 CTAAATCTGTAATTTAGTGAACTTGGC 3 |
| N7 | LOC_Os03g07300 | F 5 ATGGCGTCGCCGTCGTCGTC 3 |
| | | R 5 TCATGCTGGATCAGGCACAGG 3 |
| N9 [ox] | LOC_Os03g20460 | F 5 ATGCTTCTCACGCGAAGGTTC 3 |
| | | R 5 CTAGATATTAACCGTACGTTTGTG 3 |
| N10K | LOC_Os03g26460 | F 5 ATGAGTCGCCACCCGAGC 3 |
| | | R 5 TCACGGCTTCGCTTCACCAGC 3 |
| N11K | LOC_Os03g53720 | F 5 ATGGCGGAGGCAAGGGCGACG 3 |
| | | R 5 TTAAGTTTTGTATAGATATGAAATGTCTTGGAA 3 |
| N13 | LOC_Os04g35700 | F 5 ATGCCCGCGTGGTGGCCACG 3 |
| | | R 5 TCAGAACCTCTTTGGGGAGCC 3 |
| N15 | LOC_Os04g58850 | F 5 ATGGGTTCGGCGAGCCGCG 3 |
| | | R 5 TCAGAACCAGACTTTGCAGTC 3 |
| N16 | LOC_Os05g27730 | F 5 ATGGCGTCCTCGACGGGG 3 |
| | | R 5 CTAGCAGAGGAGCGACTCGAC 3 |
| N17K [ox] | LOC_Os06g12530 | F 5 ATGGCGATCATCTCCGACTTC 3 |
| | | R 5 CTAAGCTATTTTTGGATTGGAGAAGTC 3 |
| N19 | LOC_Os07g48290 | F 5 ATGATGGTTTCTTCTCAGACA 3 |
| | | R 5 CTAGTCTTTCAGCCTAGAGGA 3 |
| N31 | LOC_Os02g05480 | F 5 ATGTCGGCGGACGAGCTGCGG 3 |
| | | R 5 TCAAGGTAATCCCCTCACTGA 3 |
| N33 | LOC_Os08g27070 | F 5 ATGAGTCGGCACCCTGAAGTG 3 |
| | | R 5 CTAACTTGTCTCGGCTTTAGCGTC 3 |
| N38 | LOC_Os02g05480 | F 5 ATGGCGATCATGGTGGATCCT 3 |
| | | R 5 TCATCGGGCACTCATTGCTGC 3 |
| N2p | LOC_Os01g70580 | F 5 ATGCCGCTGCCGACGATGACC 3 |
| | | R 5 TTCAGAGCCAGGTGGATTTCT 3 |
| N3p | LOC_Os01g70790 | F 5 CCA GCC GCC GCG GGG C 3 |
| | | R 5 CTAGAAATCGAAGCCACCGCC 3 |
| N6p | LOC_Os02g21510 | F 5 ATGGCCGACGCTAGGTCCGC 3 |
| | | R 5 AAC ATC AAC GCC AGT CAG AGA 3 |
| N9p | LOC_Os03g20460 | F 5 ATGCTTCTCACGCGAAGGTTC 3 |
| | | R 5 GAA ATG CAT CCC CTC CCT CTC 3 |

[a] represents 238-2289 nt fragment of XB2 (failed to clone full length cDNA due to high GC content in N-terminal region).

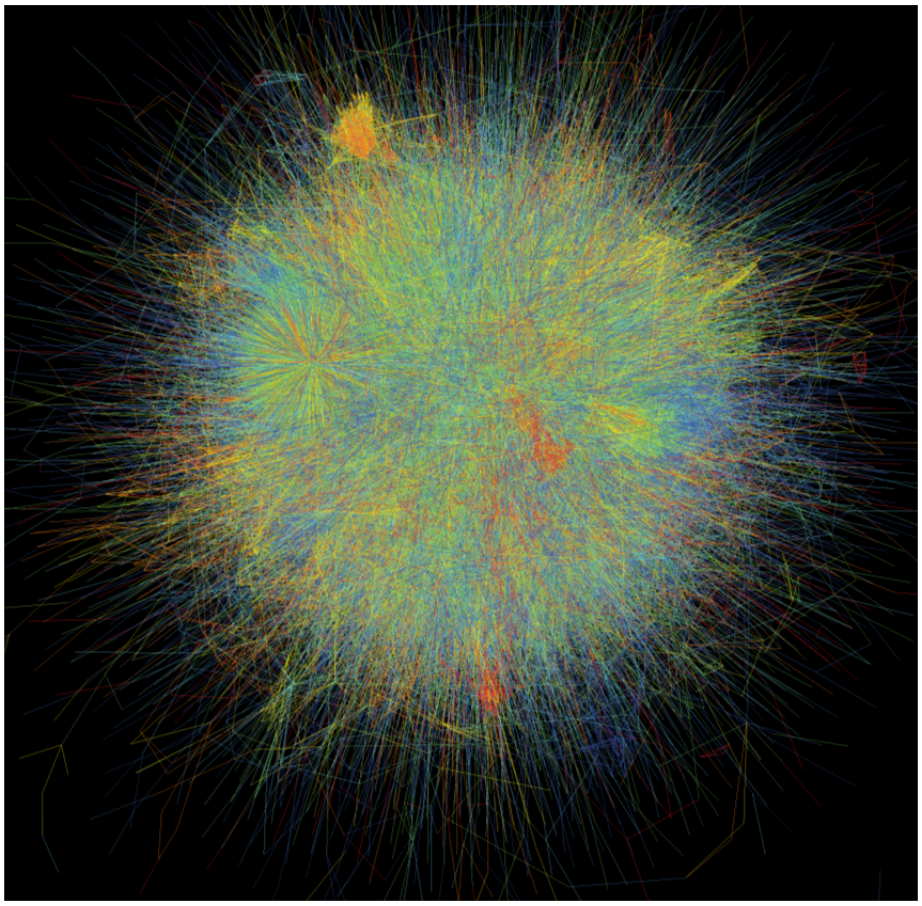
[k]cDNA fragment was amplified from a KOME (Knowledge-based Oryza Molecular biological Encyclopedia) full length cDNA (19).

[p] primers used to amplify 350-450 nt cDNA fragment to generate silencing (RNAi) construct.

[ox] primers used to amplify construct for overexpression.

**Supplementary Figure 1. An overview of the complete RiceNet, showing genes as nodes and gene-gene associations as edges.** Colors indicate the likelihood scores of gene-gene associations, with red for higher and blue for lower LLS scores. The diagram was generated by Large Graph Layout (36).

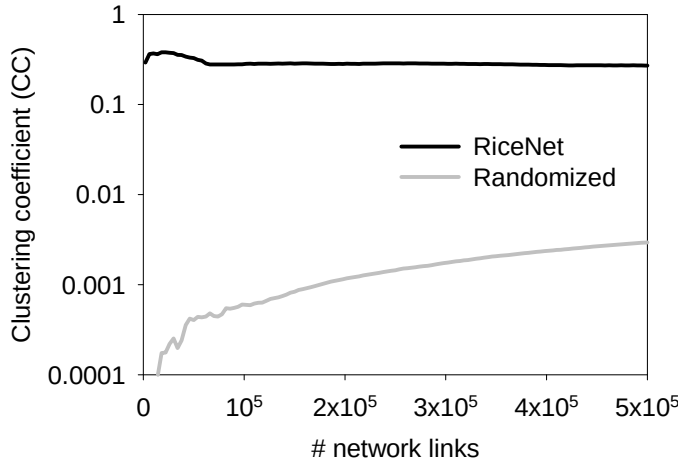
**Supplementary Figure 1.**



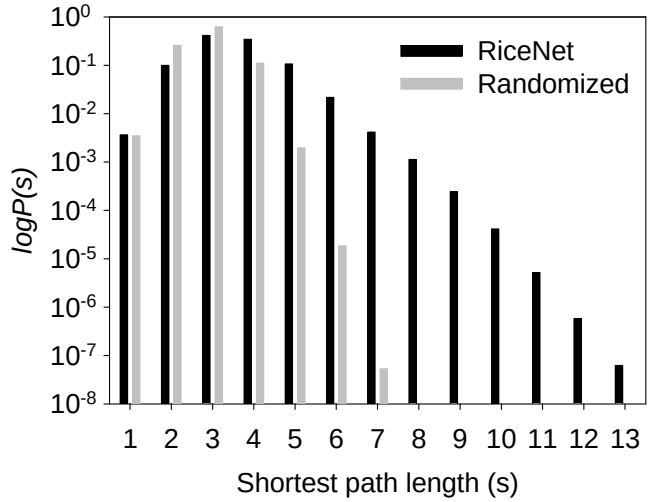
**Supplementary Figure 2. RiceNet shows a high degree of modularity, consistent with a network capturing diverse biological processes and pathways. (a)** RiceNet shows a significantly higher degree of clustering of genes, measured as the clustering coefficient (37), than does a randomized network. **(b)** RiceNet shows a significantly non-random distribution of shortest path lengths between gene pairs, indicating the presence of many regional structures (functional gene modules) separated from one another in the network.

**Supplementary Figure 2.**

**a.**

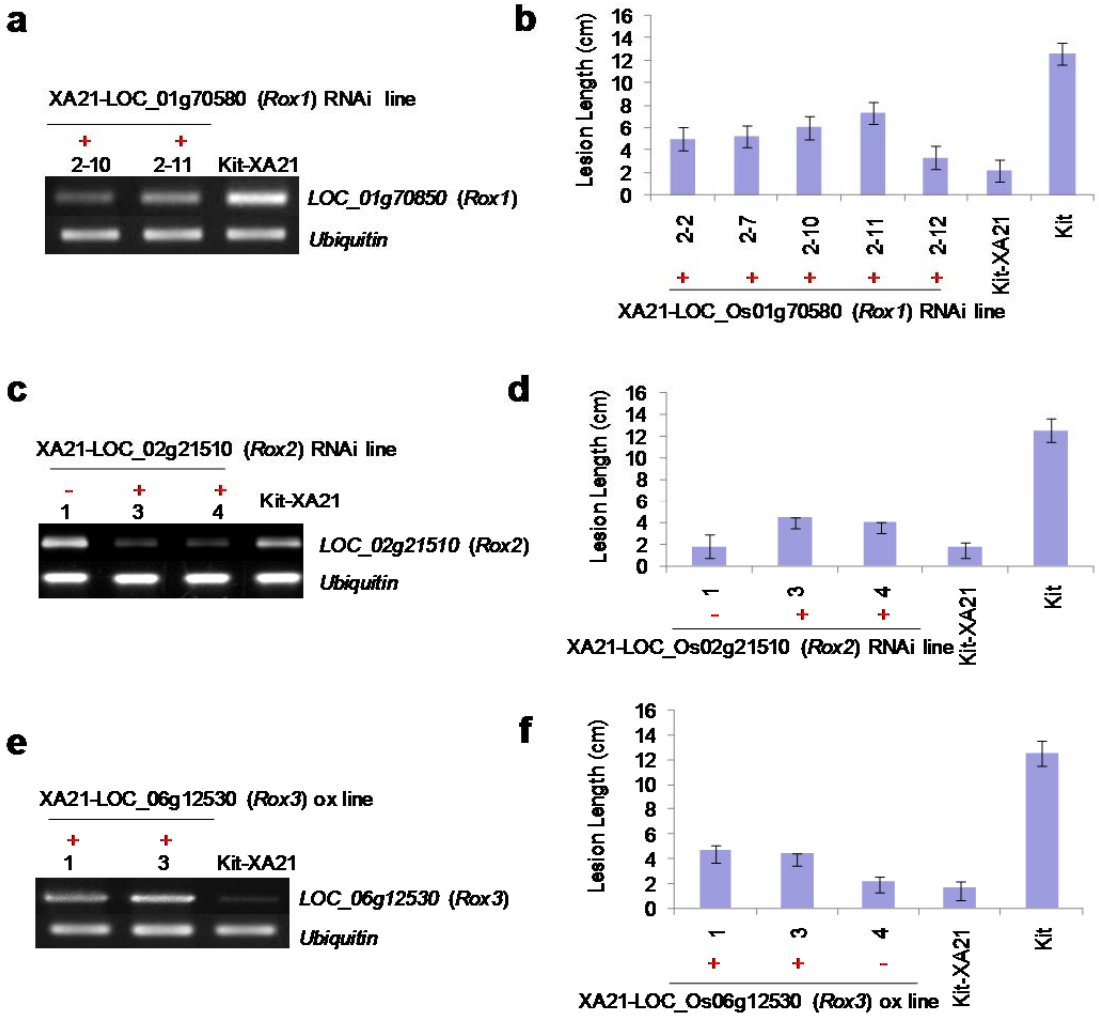



**b.**



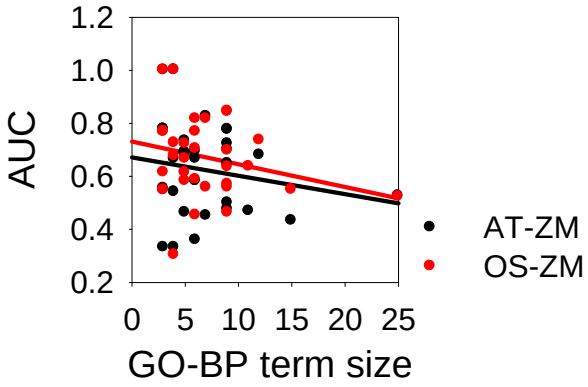
**Supplementary Figure 3. RNA quantification and phenotypes of transgenic lines altered for expression of LOC_Os01g70580 (*Rox1*), LOC_Os02g21510 (*Rox2*), and LOC_Os06g12530 (*Rox3*).** Quantification of mRNA accumulation for **(a)** LOC_Os01g70580 (*Rox1*) in T1 progeny of XA21-LOC_Os01g70580 (*Rox1*) RNAi 2-10 and 2-11 independently transformed lines, **(c)** LOC_Os02g21510 (*Rox2*) mRNA in T0 progeny of XA21-LOC_Os02g21510 (*Rox2*) RNAi 1, 3, and 4 independently transformed lines, **(e)** LOC_Os06g12530 (*Rox3)* mRNA in T0 progeny of XA21-LOC_Os06g12530 (*Rox3*) RNAi 1 and 3 independently transformed lines, along with Kitaake-XA21 (Kit-XA21) controls. Leaf lesion lengths measured after *Xoo* challenge for **(b)** T1 progeny of XA21-LOC_Os01g70580 (*Rox1*) RNAi 2-2, 2-7, 2-10, 2-11 and 2-12 lines, **(d)** T0 progeny of XA21-LOC_Os02g21510 (*Rox2*) RNAi 1, 3 and 4 lines, **(f)** T0 progeny of XA21-LOC_Os06g12530 (*Rox3*) RNAi 1, 3 and 4 lines. RT-PCR was performed using specific primers for each tested candidate gene. Ubiquitin mRNA was used as an internal control. A (-) symbol indicates lack of the transgene, and a (+) symbol indicates the presence of the transgene. Leaves were inoculated when the plants were 5 weeks old and lesion lengths were measured at 14 days after inoculation. Each bar represents the average and standard deviation from three tested leaves.

**Supplementary Figure 3.**



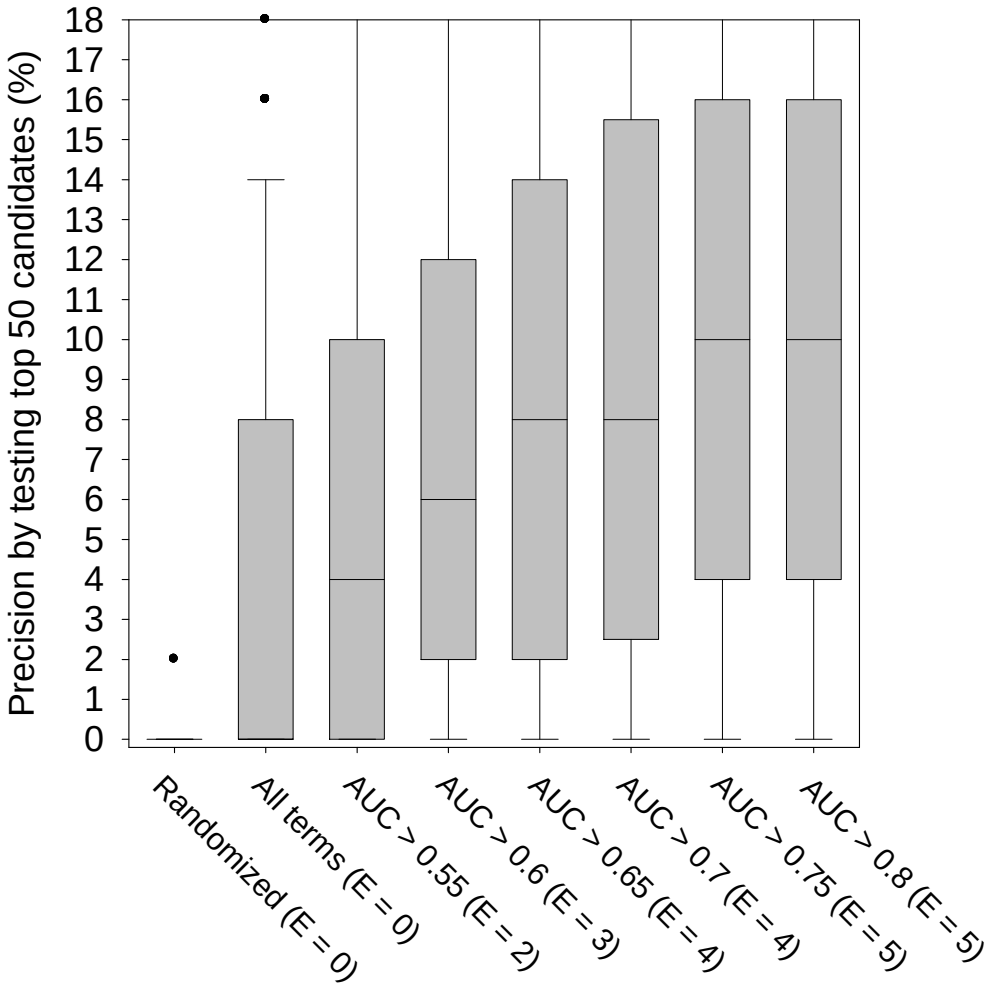
**Supplementary Figure 4. The prediction strength for different maize biological processes is largely independent of the number of genes in the process,** shown here for 32 GO biological processes for the AT-MZ and OS-MZ networks ($r^2$ = 0.03 and $r^2$ = 0.06 for AT-MZ and OS-MZ, respectively).

**Supplementary Figure 4.**



**Supplementary Figure 5. Estimates of the accuracy of predicted candidate genes as a function of AUC scores.** We calculated the prediction precision for the top 50 candidate genes for each GO biological process with >5 genes, then tallied the results as a function of prediction AUC scores. This analysis gives a rough guide to the minimum expectation for how many candidate genes are likely to be proven correct. The median precision (E) across the biological processes suggests that one might expect at least ~5 new genes among the top 50 candidates for predictions with AUC >0.75. Note that this analysis tends to underestimate precision as it only considers genes of known function in the set of positive predictions; thus, correct predictions of new candidates are penalized in this analysis.

**Supplementary Figure 5.**



**References**


1. Lee I, Ambaru B, Thakkar P, Marcotte EM, Rhee SY (2010) Rational association of genes with traits using a genome-scale gene network for Arabidopsis thaliana. *Nat Biotechnol* 28(2):149-156.
2. Lee I, Date SV, Adai AT, Marcotte EM (2004) A probabilistic functional network of yeast genes. *Science* 306(5701):1555-1558.
3. Lee I, *et al.* (2008) A single gene network accurately predicts phenotypic effects of gene perturbation in Caenorhabditis elegans. *Nat Genet* 40(2):181-188.
4. Ouyang S, *et al.* (2007) The TIGR Rice Genome Annotation Resource: improvements and new features. *Nucleic Acids Res* 35(Database issue):D883-887.
5. Kanehisa M, Goto S, Kawashima S, Nakaya A (2002) The KEGG databases at GenomeNet. *Nucleic Acids Res* 30(1):42-46.
6. Lee I, Li Z, Marcotte EM (2007) An Improved, Bias-Reduced Probabilistic Functional Gene Network of Baker's Yeast, Saccharomyces cerevisiae. *PLoS ONE* 2(10):e988.
7. Barrett T, *et al.* (2009) NCBI GEO: archive for high-throughput functional genomic data. *Nucleic Acids Res* 37(Database issue):D885-890.
8. Huynen M, Snel B, Lathe W, 3rd, Bork P (2000) Predicting protein function by genomic context: quantitative evaluation and qualitative inferences. *Genome Res* 10(8):1204-1210.
9. Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D, Yeates TO (1999) Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc Natl Acad Sci U S A* 96(8):4285-4288.
10. Wolf YI, Rogozin IB, Kondrashov AS, Koonin EV (2001) Genome alignment, evolution of prokaryotic genome organization, and prediction of gene function using genomic context. *Genome Res* 11(3):356-372.
11. Bowers PM, *et al.* (2004) Prolinks: a database of protein functional linkages derived from coevolution. *Genome Biol* 5(5):R35.
12. Dandekar T, Snel B, Huynen M, Bork P (1998) Conservation of gene order: a fingerprint of proteins that physically interact. *Trends Biochem Sci* 23(9):324-328.
13. Overbeek R, Fonstein M, D'Souza M, Pusch GD, Maltsev N (1999) The use of gene clusters to infer functional coupling. *Proc Natl Acad Sci U S A* 96(6):2896-2901.
14. Remm M, Storm CE, Sonnhammer EL (2001) Automatic clustering of orthologs

and in-paralogs from pairwise species comparisons. *J Mol Biol* 314(5):1041-1052.

15. Lee I, Blom UM, Wang PI, Shim JE, Marcotte EM (2011) Prioritizing candidate disease genes by network-based boosting of genome-wide association data. *Genome Res* 21(7):1109-1121.
16. Mostafavi S, Ray D, Warde-Farley D, Grouios C, Morris Q (2008) GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function. *Genome Biol* 9 Suppl 1:S4.
17. Wang PI, Marcotte EM (2010) It's the machine that matters: Predicting gene function and phenotype from protein networks. *J Proteomics* 73(11):2277-2289.
18. Jiao Y, *et al.* (2009) A transcriptome atlas of rice cell types uncovers cellular, functional and developmental hierarchies. *Nat Genet* 41(2):258-263.
19. Kikuchi S, *et al.* (2003) Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice. *Science* 301(5631):376-379.
20. 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(6):511-520.
21. Seo YS, *et al.* (2011) Towards establishment of a rice stress response interactome. *PLoS Genet* 7(4):e1002020.
22. Chern MS, *et al.* (2001) Evidence for a disease-resistance pathway in rice similar to the NPR1-mediated signaling pathway in Arabidopsis. *Plant J* 27(2):101-113.
23. Song WY, *et al.* (1995) A receptor kinase-like protein encoded by the rice disease resistance gene, Xa21. *Science* 270(5243):1804-1806.
24. Walia H, *et al.* (2007) Genome-wide transcriptional analysis of salinity stressed japonica and indica rice genotypes during panicle initiation stage. *Plant Mol Biol* 63(5):609-623.
25. Jain M, *et al.* (2007) F-box proteins in rice. Genome-wide analysis, classification, temporal and spatial gene expression during panicle and seed development, and regulation by light and abiotic stress. *Plant Physiol* 143(4):1467-1483.
26. Jain M, Khurana JP (2009) Transcript profiling reveals diverse roles of auxin-responsive genes during reproductive development and abiotic stress in rice. *FEBS J* 276(11):3148-3162.
27. Cheng C, *et al.* (2007) An early response regulatory cluster induced by low

temperature and hydrogen peroxide in seedlings of chilling-tolerant japonica rice. *BMC Genomics* 8:175.

28. Li M, Xu W, Yang W, Kong Z, Xue Y (2007) Genome-wide gene expression profiling reveals conserved and novel molecular functions of the stigma in rice. *Plant Physiol* 144(4):1797-1812.
29. Swarbrick PJ, *et al.* (2008) Global patterns of gene expression in rice cultivars undergoing a susceptible or resistant interaction with the parasitic plant Striga hermonthica. *New Phytol* 179(2):515-529.
30. Cho K, *et al.* (2008) Integrated transcriptomics, proteomics, and metabolomics analyses to survey ozone responses in the leaves of rice seedling. *J Proteome Res* 7(7):2980-2998.
31. Wang HJ, *et al.* (2007) Transcriptomic adaptations in rice suspension cells under sucrose starvation. *Plant Mol Biol* 63(4):441-463.
32. Hirose N, Makita N, Kojima M, Kamada-Nobusada T, Sakakibara H (2007) Overexpression of a type-A response regulator alters rice morphology and cytokinin metabolism. *Plant Cell Physiol* 48(3):523-539.
33. Mosquera G, Giraldo MC, Khang CH, Coughlan S, Valent B (2009) Interaction transcriptome analysis identifies Magnaporthe oryzae BAS1-4 as Biotrophy-associated secreted proteins in rice blast disease. *Plant Cell* 21(4):1273-1290.
34. Kato T, *et al.* (2009) Differential responses of rice to inoculation with wild-type and non-pathogenic mutants of Magnaporthe oryzae. *Plant Mol Biol* 70(6):617-625.
35. Kojima M, *et al.* (2009) Highly sensitive and high-throughput analysis of plant hormones using MS-probe modification and liquid chromatography-tandem mass spectrometry: an application for hormone profiling in Oryza sativa. *Plant Cell Physiol* 50(7):1201-1214.
36. Adai AT, Date SV, Wieland S, Marcotte EM (2004) LGL: creating a map of protein function with an algorithm for visualizing very large biological networks. *J Mol Biol* 340(1):179-190.
37. Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. *Nature* 393(6684):440-442.