Available online at www.sciencedirect.com
ScienceDirect
www.sciencedirect.com/science/journal/16720229
GENOMICS PROTEOMICS & BIOINFORMATICS

Article

Computational Identification of Protein-Protein Interactions in Rice Based on the Predicted Rice Interactome Network

Pengcheng Zhu[1#], Haibin Gu[1#], Yinming Jiao[1], Donglin Huang[1,2], and Ming Chen[1,2,3*]

[1]*Department of Bioinformatics, College of Life Sciences, Zhejiang University, Hangzhou 310058, China;*
[2]*James D. Watson Institute of Genome Sciences, Zhejiang University, Hangzhou 310058, China;*
[3]*State Key Laboratory of Plant Physiology and Biochemistry, College of Life Sciences, Zhejiang University, Hangzhou 310058, China.*

Genomics Proteomics Bioinformatics 2011 Oct; 9(4-5): 128-137. DOI: 10.1016/S1672-0229(11)60016-8

Received: Feb 23, 2011; Accepted: Jul 04, 2011

Abstract

Plant protein-protein interaction networks have not been identified by large-scale experiments. In order to better understand the protein interactions in rice, the Predicted Rice Interactome Network (PRIN; http://bis.zju.edu.cn/prin/) presented 76,585 predicted interactions involving 5,049 rice proteins. After mapping genomic features of rice (GO annotation, subcellular localization prediction, and gene expression), we found that a well-annotated and biologically significant network is rich enough to capture many significant functional linkages within higher-order biological systems, such as pathways and biological processes. Furthermore, we took MADS-box domain-containing proteins and circadian rhythm signaling pathways as examples to demonstrate that functional protein complexes and biological pathways could be effectively expanded in our predicted network. The expanded molecular network in PRIN has considerably improved the capability of these analyses to integrate existing knowledge and provide novel insights into the function and coordination of genes and gene networks.

Key words: protein-protein interactions, rice interactome, interolog, sub-network expansion, pathway clustering

Introduction

As the main carriers of biological functions, proteins seldom work solely but often require interactions with other proteins to perform their biological functions. Protein-protein interactions are present in almost every biological process in living organisms, such as DNA replication and transcription, enzyme controlled metabolic reactions, signalling transduction, protein transport, protein degradation, and cell cycle regulation. Therefore, identifying protein-protein interaction network will be highly valuable for understanding biological processes.

As far as we know, plant protein-protein interaction networks have not been identified by large-scale experiments due to the complexity of plant materials. It was reported that computational prediction of protein-protein interaction network has been performed in the dicotyledonous model plant *Arabidopsis thaliana* (*1-3*). For instance, Geisler-Lee *et al* (*2*) predicted 20,000 Arabidopsis interactions (interologs) based on homologous interactions in other species. However, computational identification of the interac-

# Equal contribution.
*Corresponding author.
E-mail: mchen@zju.edu.cn

tome is still absent in rice (*Oryza sativa*), which is an important monocotyledonous model plant and cereal crop. Currently, there are only hundreds of protein-protein interaction pairs available in rice, which are scattered across several public protein-protein interaction databases, such as IntAct (*4*).

Homologous proteins among different species have evolutionary conservation in sequences, function and structure, thus protein-protein interactions have been considered to be evolutionary conserved (*2*). Computational methods based on evolutionary conservation of protein-protein interactions in different species were known as "interolog" (*5, 6*). Interolog methods have shown advantages in prediction of protein-protein interactions associated with fundamental biological processes, which are considered as most evolutionary conserved interactions across different species. These methods have been applied in many model organisms such as human, fruit fly and Arabidopsis (*2, 7, 8*).

Recently, we reported the construction of Predicted Rice Interactome Network (PRIN), a genome-scale protein-protein interaction network in rice (*9*) using InParanoid algorithm (*10*) based on interolog method. By re-integrating protein interaction databases from six model organisms, PRIN is the most complete rice interactome database up to date (publicly available at http://bis.zju.edu.cn/prin) inferred from multiple indirect lines of evidence, including co-expression, co-localization, co-evolution, annotation similarity, domain interaction, and homologous interactions in other species. PRIN integrated 533,927 interactions with 48,152 proteins from six model organisms and identified 76,585 predicted interactions with 5,049 rice proteins (*9*). Furthermore, genomic features of rice, such as Gene Ontology (GO) annotation, subcellular localization prediction, and gene expression, were also mapped to our result for constructing a well-annotated and biologically significant network. In this study, we showed evidence that PRIN is rich enough to extract proteins with high connectivity such as ubiquitin family proteins and conserved protein-protein interactions involving evolutionarily functional proteins. In addition, we also investigated the distribution of biological pathways in PRIN and expanded some experimental sub-networks and known biological pathways.

Results and Discussion

In our network, 4,277 proteins were highly annotated by Gramene and GO database, while 57,345 predicted interactions successfully obtained co-expression data (gene co-expression Pearson correlation coefficient score), and 14,308 interactions were annotated by the subcellular localization annotation. A well-organized web-interface has been developed for database search and network visualization, which is publicly available at http://bis.zju.edu.cn/prin/.

Essential proteins

The degree of a protein in a network is an important topological property, which indicates the number of partners it interacts with. In the network we predicted, most proteins have small degree, although a few proteins interact with huge number of proteins. Thus, the network has a good fault-tolerant rate to random mutation, in order to maintain the stability of the entire network. Essential proteins always have high degree to form hubs in protein-protein interaction networks. Non-essential proteins tend to be aloof from hubs and dispersed in periphery region of the network (*7-9*). As a result, essential proteins appear to be the pivot in protein-protein interaction networks and probably are associated with many fundamental biological processes.

After extracting proteins with highest connectivity from the predicted network, we found that ubiquitin family proteins (LOC_Os06g46770.1, LOC_Os02g06640.1, LOC_Os05g42424.1, LOC_Os07g46660.1, LOC_Os01g68940.1, LOC_Os01g68950.1 and LOC_Os01g62244.1) tend to interact with highest number of proteins (**Table 1**). These data indicate that ubiquitin is highly conserved in eukaryotes. Ubiquitin may extensively participate in the protein degradation process, as well as the removal of transmembrane proteins such as receptors. Some other proteins also have a high number of interactions, including 26S proteosome (LOC_Os01g16190.1) and elongation factor Tu (LOC_Os03g08020.1, LOC_Os03g08010.1 and LOC_Os03g08050.1) (**Table 1**).

Most conserved interactions

The rice interactome we predicted was based on

Table 1 Top degree proteins in predicted interactome

Protein	Degree	Description
LOC_Os06g46770.1	794	ubiquitin family protein, putative, expressed
LOC_Os02g06640.1	695	ubiquitin family protein, putative, expressed
LOC_Os05g42424.1	695	ubiquitin family protein, putative, expressed
LOC_Os01g16190.1	561	26S proteosome non-ATPase regulatory subunit 14, putative, expressed
LOC_Os03g08020.1	528	elongation factor Tu, putative, expressed
LOC_Os03g08010.1	528	elongation factor Tu, putative
LOC_Os03g08050.1	528	elongation factor Tu, putative, expressed
LOC_Os09g30412.1	484	heat shock protein, putative, expressed
LOC_Os07g41180.1	460	RNA-binding protein-like, putative, expressed
LOC_Os07g46660.1	459	ubiquitin carboxyl-terminal hydrolase domain containing protein, expressed
LOC_Os01g73310.1	434	actin, putative, expressed
LOC_Os06g37180.1	381	ATP synthase, putative, expressed
LOC_Os03g42900.1	377	KH domain containing protein, putative, expressed
LOC_Os01g68940.1	367	ubiquitin family domain containing protein, expressed
LOC_Os10g32550.1	359	T-complex protein, putative, expressed
LOC_Os01g62840.1	346	mannose-1-phosphate guanyltransferase, putative, expressed
LOC_Os07g08330.1	342	ribosomal protein L4, putative, expressed
LOC_Os01g68950.1	324	ubiquitin family domain containing protein, expressed
LOC_Os01g62244.1	306	ubiquitin-conjugating enzyme, putative, expressed
LOC_Os07g31370.1	285	ras-related protein, putative, expressed

conserved protein-protein interactions among multi-species. As reported by previous studies, many fundamental pathways show their evolutionary conservation among species, such as GTPase signal transduction (*11*). We derived 20 most conserved interactions during the interolog prediction, which were sorted by co-expression Pearson correlation coefficient (PCC) score and relative specificity similarity (RSS^{GO}) score. These interactions involve regulatory protease, kinase in cell cycle, DNA repair process and RNA binding process, which are obviously associated with fundamental biological processes. It should be noted that some evolutionarily conserved interactions contain proteins with no annotation in rice proteome, such as LOC_Os02g37920.1, whose ortholog proteins in human, yeast, fruit fly and nematode were reported to interact with proteins involved in DNA mismatch repair process (*12*). Consequently, evolutionary conserved interactions can be used to identify unknown proteins.

With construction of sub-network for top conserved protein-protein interactions (**Figure 1A**), we find that evolutionarily conserved proteins tend to co-express with each other and share significant correlation in GO annotation (as given by RSS^{GO} score in **Table 2**). A degree distribution was generated by the topological analysis of evolutionarily conserved proteins (**Figure 1B**), and it was noticed that most proteins had degree higher than 50. This discovery suggests that evolutionarily conserved functional proteins have more interactions in interactome, and as mentioned above, proteins with high degrees appear to be the hub in the interactome. Therefore, we speculate that these proteins play important and fundamental roles in the rice proteome.

Expanding experimental sub-network

Among 430 experimentally determined rice protein-protein interactions, 406 proteins have been derived from BIND (*13*), IntAct (*4*) and PlaPID (*14*). Although the experimental interactome has a low coverage on the whole rice interactome, 95 proteins constituting 230 interactions in our network were found to confer 66 interactions in the small experimental network. Encouragingly, 20 of these 66 interactions have been confirmed by experiments, reflecting an appreciable sensitivity in spite of the rare experimental data.

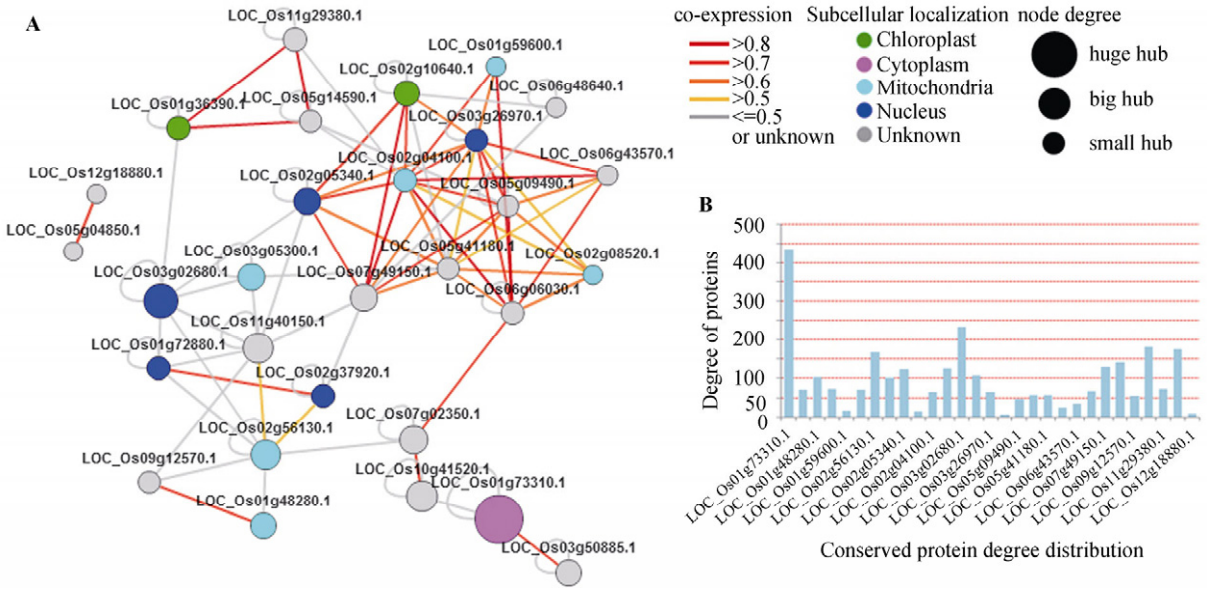

Figure 1 Evolutionarily conserved interaction sub-network. **A**. A sub-network constructed by most evolutionarily conserved proteins in Table 2. **B**. Statistics of degree distribution for conserved proteins. Degrees for most of conserved proteins are higher than 50.

Table 2 Top conserved interactions in predicted interactome.

Protein A	Protein B	Description A	Description B	PCC score	RSS^{CC}	RSS^{MF}	RSS^{BP}	Species
LOC_Os02g05340.1	LOC_Os07g49150.1	proteasome/cyclosome repeat containing protein	26S protease regulatory subunit 4	0.72	1	0.81	0.78	5
LOC_Os06g48640.1	LOC_Os07g49150.1	proteasome/cyclosome repeat containing protein	26S protease regulatory subunit 4	0.39	1	0.81	0.78	5
LOC_Os03g02680.1	LOC_Os03g05300.1	cyclin-dependent kinase A-1	cyclin-dependent kinases regulatory subunit 1			0.51		5
LOC_Os02g56130.1	LOC_Os02g56130.1	PCNA – DNA replicative polymerase clamp	PCNA – DNA replicative polymerase clamp					5
LOC_Os11g40150.1	LOC_Os11g40150.1	DNA repair protein Rad51	DNA repair protein Rad51					5
LOC_Os01g36390.1	LOC_Os05g14590.1	MCM complex subunit 4	MCM complex subunit 6	0.91	1	1	1	4
LOC_Os05g14590.1	LOC_Os11g29380.1	MCM complex subunit 6	MCM complex subunit 2	0.89	1	1	1	4
LOC_Os01g59600.1	LOC_Os02g04100.1	peptidase, T1 family	peptidase, T1 family	0.8	1	1	1	4
LOC_Os05g04850.1	LOC_Os12g18880.1	RNA recognition motif protein	mago nashi	0.76	1			4
LOC_Os02g04100.1	LOC_Os03g26970.1	T1 family protein	T1 family protein	0.75	1	1	1	4
LOC_Os03g26970.1	LOC_Os06g06030.1	T1 family protein	T1 family protein	0.75	1	1	1	4
LOC_Os02g05340.1	LOC_Os02g10640.1	proteasome/cyclosome repeat protein	26S protease regulatory subunit	0.74	1	0.81	0.78	4
LOC_Os01g72880.1	LOC_Os02g37920.1	DNA mismatch repair protein	Expressed protein	0.73		1	1	4
LOC_Os01g73310.1	LOC_Os03g50885.1	actin	actin	0.73			1	4

With the integration of experimental data, we finally got a small protein-protein interaction network of rice (**Figure 2A**). Since these rice protein-protein interactions were derived from various specific research of rice proteome, the small experimental interactome shows a low degree of connectivity and a high degree of modularization. A small number of proteins tend to gather into clusters or cliques associated with specific biological process. However, these clusters are far from complete due to the absence of large-scale experiments in rice. Take the proteins commonly found both in experimental and predicted rice network as seed proteins, the resulting sub-networks are included in our predicted network.

Here we take the sub-network of MADS-box domain-containing proteins as an example. MADS-box domain is conserved in plant, with typical length of 168 to 180 base pairs. MADS-box domain-containing proteins are essential for sequence-specific DNA binding and dimerization in plants. It has been reported in *A. thaliana* that MADS-box genes participated in the determination of floral organ identity and flowering time pathways (*15, 16*). Fifteen interactions among seven rice MADS-box containing proteins have been determined in the earlier studies by Moon *et al* (*17*). In the experimental interactome, these 7 proteins construct a highly modular sub-network (**Figure 2B**), in which 2 of 15 interactions (interaction between OsMADS6 and OSMADS14, and OSMADS14 self-interaction) are also detected in our

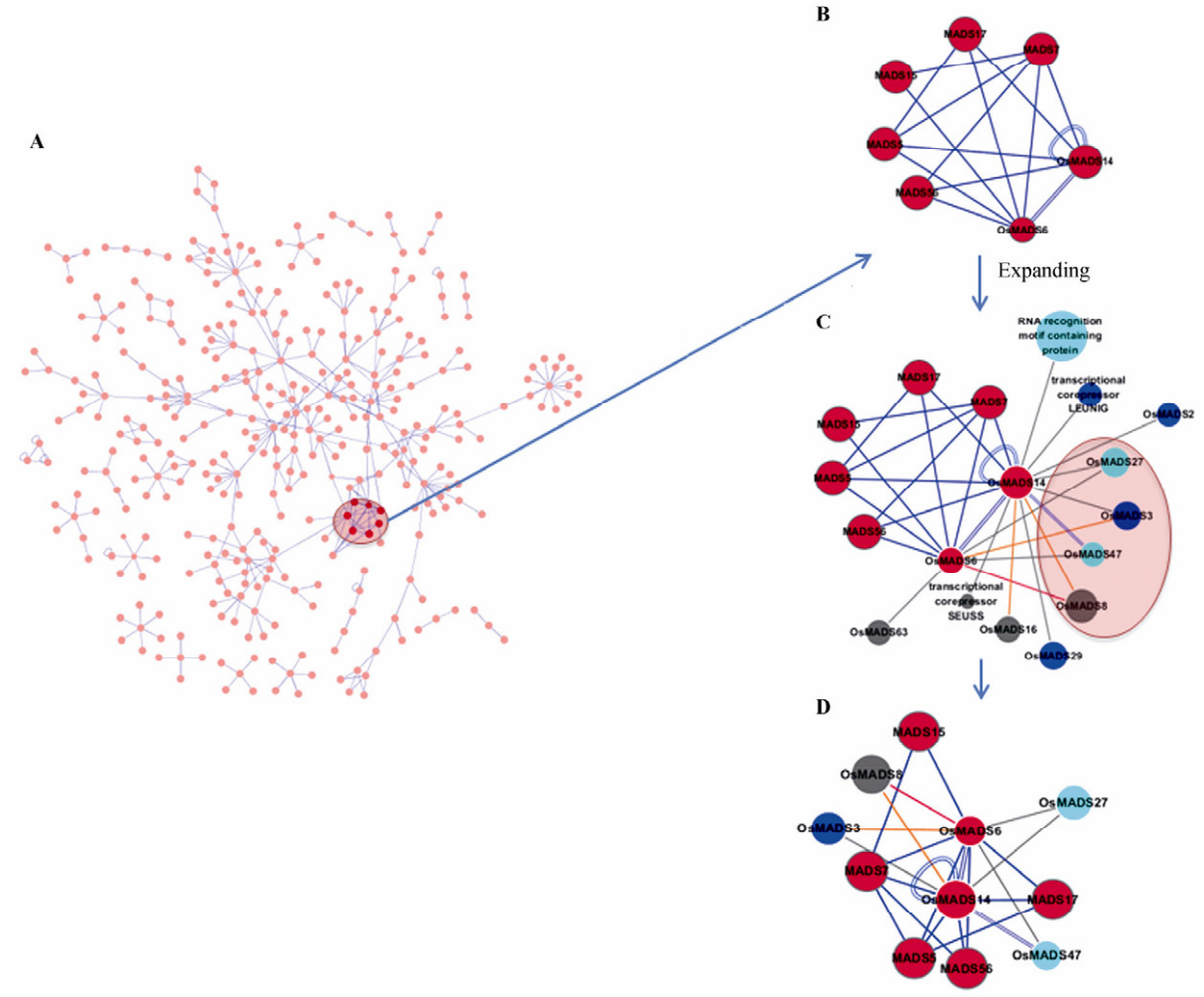

Figure 2 Expansion of MADS-box containing proteins. **A**. A small interactome network of rice with experimental identification visualized by Cytoscape. **B**. A functional motif constructed by seven MADS-box containing proteins in rice. **C**. An expanded sub-network of MADS-box related motif. Proteins in pink region are predicted as potential participants to the protein complex due to their tight interactions with the motif. **D**. A reconstructed sub-network with new participants.

predicted network. These seven proteins (OsMADS5, OsMADS6, OsMADS7, OSMADS14, OsMADS15, OsMADS17 and OsMADS56) are considered as a functional motif in the network due to their tight interrelation. In addition, two proteins, OsMADS6 and OsMADS14, are used as seed proteins with their first neighbors in our network to create an expanded sub-network for further in-depth identification. Eleven proteins have direct interactions with the seed proteins in our network, eight of which have been annotated as MADS-box containing proteins. Among these eight proteins, OsMADS3, OsMADS8, OsMADS27 and OsMADS47 have been found to interact with both OsMADS6 and OsMADS14, which are potential participants in the sub-network motif (**Figure 2C**). Moreover, the interaction between OsMADS47 and OsMAD14 was reported previously (*18*), supporting the reliability of our network. Finally, an expanded sub-network was constructed by 11 MADS-box containing proteins, among which 4 proteins are new participants (**Figure 2D**).

Biological pathway clustering

Protein-protein interactions are the basic composition of biological pathways, and play fundamental roles in almost all biological processes. We derived the rice biological pathway data from the most comprehensive biological pathway database KEGG (*19*). There are 2,235 proteins in 112 pathways derived from KEGG, among which 698 proteins are also found in our network, forming 5,010 interactions. Next we investigated the rice biological pathway distribution in our network. We take the protein coverage as a measure of pathway integrity. **Figure 3A** shows the top covered rice biological pathways in our predicted network, and the number in orange shows the number of common proteins found in our network and KEGG pathways. Sphingolipid metabolism pathway has the highest integrity in our predicted network, with eight proteins contributing 89% coverage of whole pathway. Proteasome pathway has the highest number of common proteins with a good coverage (80%) in our network. Based on the statistics of clueGO (*20*), 31 pathways have over 50% coverage in our predicted network (**Figure 3B**). We use clueGO to classify all the pathways into 11 pathway clusters except individual pathways. It appears that most proteins in our network fall into pathways relating to valine, leucine and isoleucine metabolism processes. Glycolysis/gluconeogenesis pathways come to the next and consist of 184 proteins. Androgen and estrogen related pathways, phenylpropanoid biosynthesis pathways, glutamate metabolism pathways, sphingolipid metabolism pathways and purine metabolism pathways also contain considerable numbers of proteins. Proteins associated with clustered pathways account for 71% of proteins involved in all pathways (the remaining 29% participate in the individual pathways), which indicates the modular properties of our predicted interactome.

Expanding known biological pathways

Protein-protein interaction network based on computational methods provides insights for potential functions of the proteins in various biological processes, which could guide and be validated by large-scale experiments. One of the key tasks for systems biology is biological pathway prediction. Taking function-specific proteins in the network as seed proteins to expand existing biological pathways is commonly used to effectively discover new participants of known pathways. Here we take a plant circadian rhythm related pathway as an example to demonstrate how to use our predicted network to expand known pathways.

Circadian rhythm signaling pathways is one of the most complicated signalling networks in plant physiological processes. Photosynthesis, respiration, plant nutrition and plant hormone all involve circadian rhythm signaling pathways, which adjust the rhythm on a daily basis. Light and temperature are two major periodic changes of environment, which influence plant circadian rhythm. Plants have an endogenous central oscillator that regulates many aspects of circadian rhythm, such as photoperiodic behavior. Multiple proteins are related to the oscillator and participate in the process of circadian rhythm control (*21*). Sixteen proteins have been reported as participants in plant circadian rhythm signalling pathways. Among them, seven proteins are also found in our predicted network (**Figure 4A**), which form a small sub-network with nine interactions (**Figure 4B**).

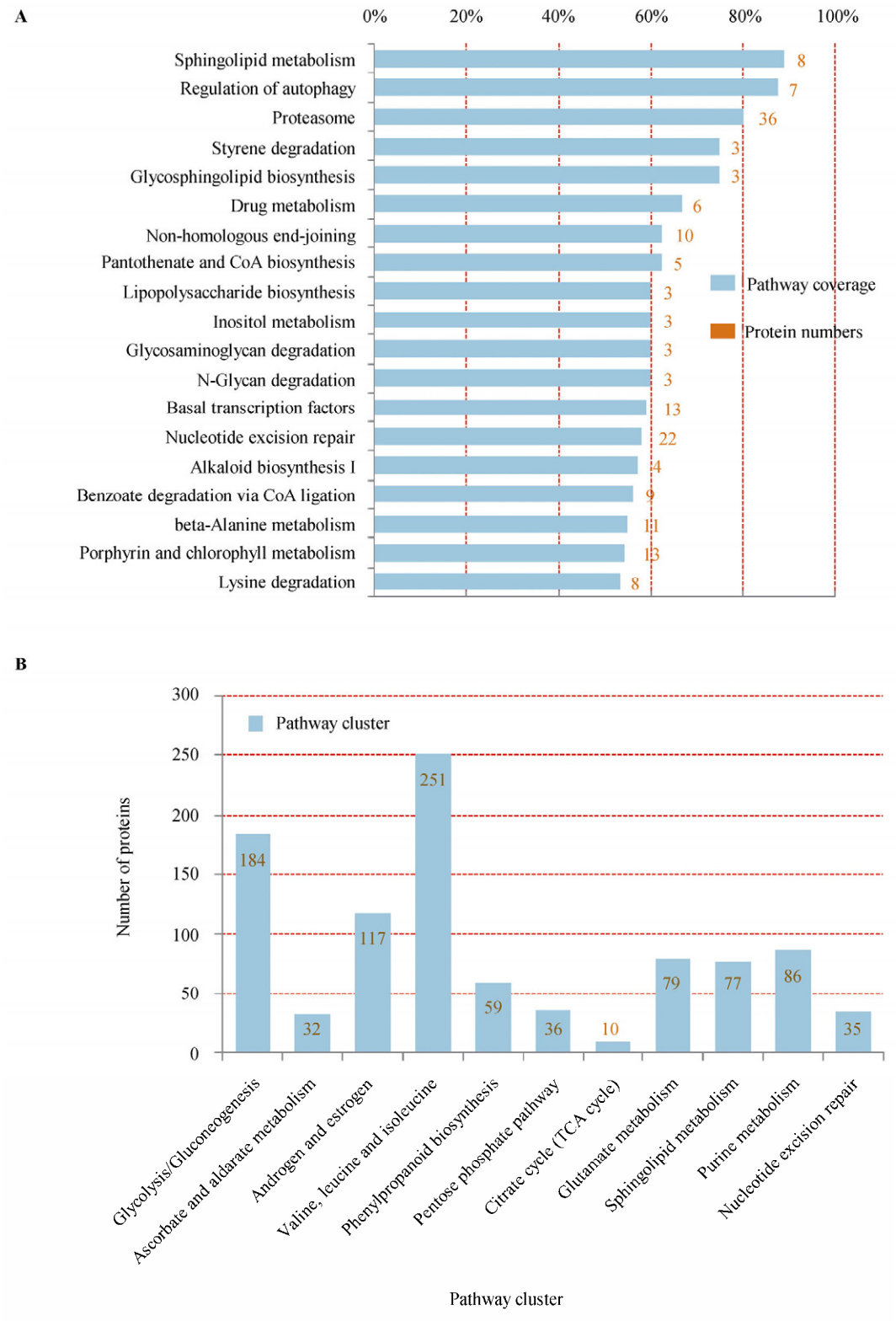

Figure 3 Mapping and clustering of KEGG pathways. **A**. Statistics of biological pathways in rice derived from KEGG, which have the highest coverage in our network. Blue bars show the coverage of pathways in our network, and numbers in orange indicate numbers of common proteins between these pathways and our network. **B**. A pathway cluster derived using clueGO.

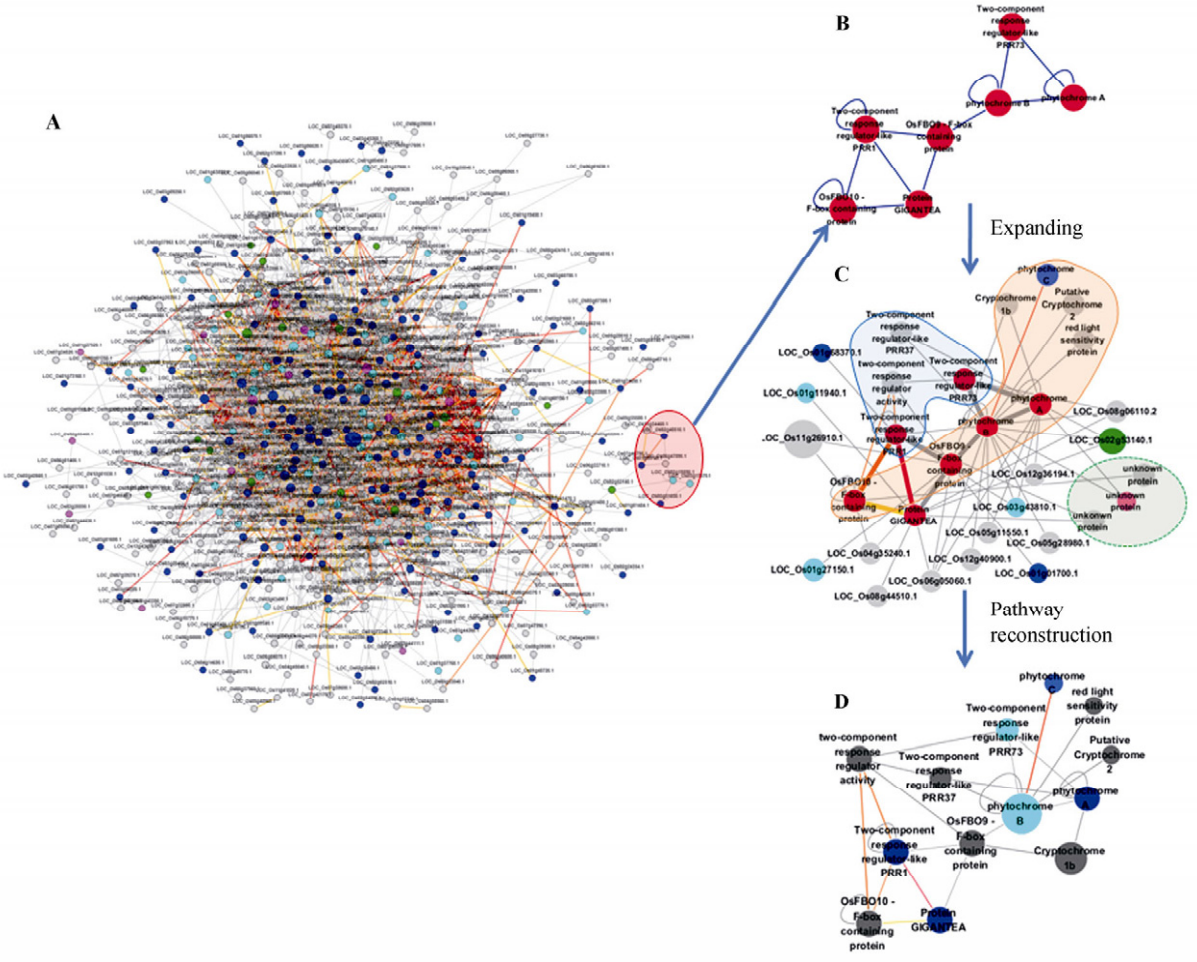

Figure 4 Expansion of circadian rhythm signaling pathways. **A**. A KEGG pathway mapped to our predicted network. **B**. Sub-network of circadian rhythm signaling pathways constructed by seven proteins. **C**. An expanded sub-network of rhythm signaling pathway-related proteins. Proteins in blue region are two-component response regulator proteins. Proteins in orange are associated with plant phototonus related process. Proteins in green have no previously known function, which have potential molecular function in rice photosynthesis based on our network prediction. **D**. The expanded circadian rhythm signaling pathways.

Four of these seven proteins are associated with plant phototonus related process (Phytochrome A, Phytochrome B, OsFBO9 and OsFBO10), while another two of them are associated with two-component response regulator activity (PRR73 and PRR1). We take these seven proteins as seed proteins and extract their first neighbors in our network to create an expanded sub-network (**Figure 4C**). Finally, 24 proteins are found to have direct interactions with the seed proteins in our network. Among these 24 proteins, 4 proteins are found associated with plant phototonus related process (LOC_Os04g37920.1, LOC_Os03g 54084.1, LOC_Os02g41550.4 and LOC_Os05g 02690.1). Notably, LOC_Os04g37920.1 and LOC_Os02g41550.4 were annotated as cryptochrome, and LOC_Os04g37920.1 has significant interactions both with Phytochrome A and OsFBO9. LOC_Os03g 54084.1, which was annotated as Phytochrome C, has significant coexpression with Phytochrome B. LOC_Os05g02690.1 was a potential participant of photosynthesis due to its sensitivity to red light. Additionally, proteins LOC_Os07g49460.1 and LOC_Os11g05930.1, which were annotated as two-component response regulator, were also found in the expanded sub-network with similar functional annotations to PRR73 and PRR1(**Figure 4D**). The former has significant interactions with Phytochrome A and Phytochrome B, while the latter has significant interactions

with PRR73, OsFBO9, PRR1 and OsFBO10. Furthermore, the latter also has significant co-expression with PRR1 and OsFBO10. Consequently, these two proteins are postulated to be potential participants of rice circadian rhythm pathways. On the other hand, proteins with no previously known function also emerged with the seed proteins (LOC_Os01g15990.1, LOC_Os07g 38360 and LOC_Os07g48570), which have potential function in rice photosynthesis based on our network prediction.

Conclusion

Using interolog of 6 model organisms, we have identified 76,585 interactions involving 5,049 rice proteins in PRIN. By extracting the most connective proteins from our predicted network, we found that ubiquitin family proteins (LOC_Os06g46770.1, LOC_Os02g06640.1, LOC_Os05g42424.1, LOC_Os07g 46660.1, LOC_Os01g68940.1, LOC_Os01g68950.1 and LOC_Os01g62244.1) tend to interact with highest number of proteins. We also derived 20 most conserved interactions during the interolog prediction, which are sorted by co-expression PCC score and RSS^{GO} score. Furthermore, the biological pathway distribution in the network was investigated in rice. It showed that most proteins in our network fall into pathways relating to valine, leucine and isoleucine metabolism processes. Plant circadian rhythm related pathway was taken as an example to demonstrate how to use our predicted network to expand known pathways. The results indicated that functional protein complexes and biological pathways could be effectively expanded in our predicted network, which will provide new insights on the protein-protein interaction network in rice.

Materials and Methods

PRIN database

The predicted protein-protein interaction network in rice, PRIN, was derived from six model species including *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, *Drosophila melanogaster*, *Homo sapiens*, *Escherichia coli* K12 and *Arabidopsis thaliana* based on the interolog method as described previously (*9*). The PRIN database is publicly available at http://bis.zju.edu.cn/prin/.

Network annotation

GO is an important functional annotation for proteome. Gene ontologies of rice were derived from GO database (*22*) and Gramene database (*23*). RSS^{GO} scores, which are mainly based on GO term similarity and GO depth (*24, 25*), are calculated for every GO annotated interaction (cell component, biological process and molecular function, respectively) using the method provided by SPIDer (*26*). The PCC scores of an interaction in our network were obtained from the RiceArray Database (*27*) calculation. The calculation is based on rice gene expression data in 830 rice Affymetrix microarray data (NCBI GEO AC: GPL2025). Subcellular localization annotations of rice proteome were obtained from RSLpred prediction (*28*), which is a specific predictor for rice.

Acknowledgements

We thank Peijian Cao, Fei He, Xiao Li, Kui Lin and Christian Klukas for their generous help. This work was supported by the National Natural Science Foundation of China (Grant No. 30771326, 30971743, 31050110121), the National Science and Technology Project of China (Grant No. 2008AA10Z125, 2008ZX08003-005, 2009DFA32030), and the Program for New Century Excellent Talents in University of China (Grant No. NCET-07-0740).

Authors' contributions

MC conceived the idea of this research. PZ designed the method, analyzed the results, and prepared the manuscript. HG constructed the database and web interface. YJ tested the web server. DH provided advice for network analysis. All authors read and approved the final manuscript.

Competing interests

The authors have declared that no competing interests exist.

References

1 Lin, M., *et al*. 2009. Computational identification of potential molecular interactions in Arabidopsis. *Plant Physiol.* 151: 34-46.

2 Geisler-Lee, J., *et al*. 2007. A predicted interactome for Arabidopsis. *Plant Physiol.* 145: 317-329.

3 Lin, M., *et al*. 2011. PAIR: the predicted Arabidopsis interactome resource. *Nucleic Acids Res.* 39: D1134-1140.

4 Kerrien, S., *et al*. 2007. IntAct—open source resource for molecular interaction data. *Nucleic Acids Res.* 35: D561-565.

5 Brown, K.R. and Jurisica, I. 2007. Unequal evolutionary conservation of human protein interactions in interologous networks. *Genome Biol.* 8: R95.

6 Matthews, L.R., *et al*. 2001. Identification of potential interaction networks using sequence-based searches for conserved protein-protein interactions or "interologs". *Genome Res*. 11: 2120-2126.

7 Huang, T.W., *et al*. 2007. Reconstruction of human protein interolog network using evolutionary conserved network. *BMC Bioinformatics* 8: 152.

8 Brown, K.R. and Jurisica, I. 2005. Online predicted human interaction database. *Bioinformatics* 21: 2076-2082.

9 Gu, H., *et al*. 2011. PRIN, a predicted rice interactome network. *BMC Bioinformatics* 12: 161.

10 O'Brien, K.P., *et al*. 2005. Inparanoid: a comprehensive database of eukaryotic orthologs. *Nucleic Acids Res.* 33: D476-480.

11 Carter, C.J., *et al*. 2004. Membrane trafficking in plants: new discoveries and approaches. *Curr. Opin. Plant Biol.* 7: 701-707.

12 Heck, J.A., *et al*. 2006. Negative epistasis between natural variants of the *Saccharomyces cerevisiae* MLH1 and PMS1 genes results in a defect in mismatch repair. *Proc. Natl. Acad. Sci. USA* 103: 3256-3261.

13 Willis, R.C. and Hogue, C.W. 2006. Searching, viewing, and visualizing data in the Biomolecular Interaction Network Database (BIND). *Curr. Protoc. Bioinformatics* Chapter 8: Unit 8.9.

14 Min, M., *et al*. 2010. PlaPID: a database of protein-protein interactions in plants. In *Proceedings of the Fourth International Conference on Bioinformatics and Biomedical Engineering*, Chengdu, China.

15 Onouchi, H., *et al*. 2000. Mutagenesis of plants overexpressing CONSTANS demonstrates novel interactions among Arabidopsis flowering-time genes. *Plant Cell* 12: 885-900.

16 Michaels, S.D. and Amasino, R.M. 1999. FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant Cell* 11: 949-956.

17 Moon, Y.H., *et al*. 1999. Determination of the motif responsible for interaction between the rice APETALA1/AGAMOUS-LIKE9 family proteins using a yeast two-hybrid system. *Plant Physiol.* 120: 1193-1204.

18 Cooper, B., *et al*. 2003. A network of rice genes associated with stress response and seed development. *Proc. Natl. Acad. Sci. USA* 100: 4945-4950.

19 Kanehisa, M. 2002. The KEGG database. *Novartis Found. Symp.* 247: 91-101; discussion 101-103, 119-128, 244-252.

20 Bindea, G., *et al*. 2009. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* 25: 1091-1093.

21 McWatters, H.G., *et al*. 2001. Picking out parallels: plant circadian clocks in context. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 356: 1735-1743.

22 Harris, M.A., *et al*. 2004. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* 32: D258-261.

23 Youens-Clark, K., *et al*. 2010. Gramene database in 2010: updates and extensions. *Nucleic Acids Res.* 39: D1085-1094.

24 Wu, H., *et al*. 2005. Prediction of functional modules based on comparative genome analysis and Gene Ontology application. *Nucleic Acids Res.* 33: 2822-2837.

25 Wu, X., *et al*. 2006. Prediction of yeast protein-protein interaction network: insights from the Gene Ontology and annotations. *Nucleic Acids Res*. 34: 2137-2150.

26 Wu, X., *et al*. 2006. SPIDer: Saccharomyces protein-protein interaction database. *BMC Bioinformatics* 7: S16.

27 Jung, K.H., *et al*. 2008. Refinement of light-responsive transcript lists using rice oligonucleotide arrays: evaluation of gene-redundancy. *PLoS One* 3: e3337.

28 Kaundal, R. and Raghava, G.P. 2009. RSLpred: an integrative system for predicting subcellular localization of rice proteins combining compositional and evolutionary information. *Proteomics* 9: 2324-2342.