



# Predicting Protein–Protein Interactions Between Rice and Blast Fungus Using Structure-Based Approaches

Cunjian Zheng, Yuan Liu, Fangnan Sun, Lingxia Zhao\* and Lida Zhang\*

Department of Plant Science, School of Agriculture and Biology, Shanghai Jiao Tong University, Shanghai, China

## OPEN ACCESS

### Edited by:

Raul Antonio Sperotto,  
Universidade do Vale do  
Taquari—Univates, Brazil

### Reviewed by:

Vincent Were,  
The Sainsbury Laboratory,  
United Kingdom  
Santosh Kumar Gupta,  
National Institute of Plant Genome  
Research (NIPGR), India  
Joshitha Vijayan,  
ICAR—National Rice Research  
Institute, India

### \*Correspondence:

Lida Zhang  
zhangld@sjtu.edu.cn  
Lingxia Zhao  
lxzhao@sjtu.edu.cn

### Specialty section:

This article was submitted to  
Plant Biophysics and Modeling,  
a section of the journal  
Frontiers in Plant Science

**Received:** 12 April 2021

**Accepted:** 21 June 2021

**Published:** 23 July 2021

### Citation:

Zheng C, Liu Y, Sun F, Zhao L and  
Zhang L (2021) Predicting  
Protein–Protein Interactions Between  
Rice and Blast Fungus Using  
Structure-Based Approaches.  
*Front. Plant Sci.* 12:690124.  
doi: 10.3389/fpls.2021.690124

Rice blast, caused by the fungus *Magnaporthe oryzae*, is the most devastating disease affecting rice production. Identification of protein–protein interactions (PPIs) is a critical step toward understanding the molecular mechanisms underlying resistance to blast fungus in rice. In this study, we presented a computational framework for predicting plant–pathogen PPIs based on structural information. Compared with the sequence-based methods, the structure-based approach showed to be more powerful in discovering new PPIs between plants and pathogens. Using the structure-based method, we generated a global PPI network consisted of 2,018 interacting protein pairs involving 1,344 rice proteins and 418 blast fungus proteins. The network analysis showed that blast resistance genes were enriched in the PPI network. The network-based prediction enabled systematic discovery of new blast resistance genes in rice. The network provided a global map to help accelerate the identification of blast resistance genes and advance our understanding of plant–pathogen interactions.

**Keywords:** protein-protein interactions, rice, blast fungus (*Magnaporthe oryzae*), protein structure, machine learning

## INTRODUCTION

Rice blast, caused by the fungus *Magnaporthe oryzae*, is the most devastating disease affecting rice production. Due to the availability of both genome sequences and the accessibility of molecular genetic tools, the pathosystem between rice and blast fungus has become a model system for studying plant–pathogen interactions (Dean et al., 2005; International Rice Genome Sequencing Project, 2005). Although the molecular mechanisms of the plant immune system have been extensively investigated over the past decade, many aspects of the overall resistance picture remain poorly understood (Meng et al., 2019).

Protein–protein interactions (PPIs) play a critical role in molecular recognition between plants and pathogens. Identification of these PPIs is important for understanding the underlying molecular mechanisms against pathogen infection in plants. Experimental methods have been used to identify plant–pathogen PPIs (Mukhtar et al., 2011; Weßling et al., 2014; Cao et al., 2019), but the available interaction data are still far from depicting global maps of plant–pathogen interactions (Ammari et al., 2016). Only a few experimentally verified PPIs between rice and blast fungus have been reported in the individual studies, which is insufficient to elucidate the molecular mechanisms leading to disease resistance in rice (Kanzaki et al., 2012; Cesari et al., 2013).

To complement experimental methods for identifying PPIs, many computational methods have been developed to accelerate the discovery of PPIs (Tanwar and George Priya Doss, 2018). Most available computational methods such as interolog mapping (Matthews et al., 2001), domain-based inference (Deng et al., 2002), gene fusion (Marcotte et al., 1999), phylogenetic similarity (Pellegrini et al., 1999), and gene co-expression [28] are primarily focused on determining PPIs within a single organism. Some of these methods, such as interolog and domain-based inference, have also been applied to the interspecies PPI field and lead to the discovery of important biological insights in plant–pathogen interactions (Li et al., 2012; Sahu et al., 2014; Yang et al., 2019).

Recently, computational methods using structural information to predict PPIs have gained much attention due to the rapid growth of protein three-dimensional (3D) structures (Zhang et al., 2012; Burley et al., 2017). Rather than sequence-based methods, structure-based approaches could reveal the structural details of protein interactions (Mariano and Wuchty, 2017). However, the structure-based methods require the 3D characterization of each protein and therefore suffer from low coverage of the proteome. Predicting protein interactions based on homology-modeled structures might be a solution to this problem, which enabled the use of protein structural information on a genome-wide scale (Zhang et al., 2016; Liu et al., 2017; Zhao et al., 2019).

In this study, we presented a computational framework for predicting plant–pathogen PPIs based on structural information. Performance assessment showed that the structure-based method was powerful in discovering PPIs between plants and pathogens. Furthermore, we used the structure-based method to generate a global PPI network between rice and blast fungus proteins, which provided a valuable reference for systems understanding of plant and pathogen interactions.

## MATERIALS AND METHODS

### Data Sources

All experimentally verified host–pathogen interactions were collected from the Host–Pathogen Interaction Database (<https://hpidb.igbb.msstate.edu/index.html>) (Ammari et al., 2016). A total of 10,148 host–pathogen PPIs with homology models and experimental structures were used as the positive reference dataset, while the interacting protein pairs between hosts and pathogens were randomly shuffled to form the negative reference dataset.

The genome of *M. oryzae* strain 70–15 was downloaded from Ensembl Genomes ([http://fungi.ensembl.org/Magnaporthe\\_oryzae/Info/Index](http://fungi.ensembl.org/Magnaporthe_oryzae/Info/Index)), and the genome of *Oryza sativa* ssp. *japonica* cv. Nipponbare was downloaded from the MSU Rice Genome Annotation Project Database (<http://rice.plantbiology.msu.edu/>). A total of 38,864 non-transposable element protein sequences were identified in the rice genome.

### Identification of Membrane and Secreted Proteins in Blast Fungus

The membrane proteins containing one or more transmembrane helices were predicted by using TMHMM (Möller et al., 2001). The putatively secreted proteins were identified when the protein containing a signal peptide was predicted by using SignalP-5.0 (Almagro Armenteros et al., 2019) and the extracellular localization of the protein was predicted by using WoLF PSORT (Horton et al., 2007).

### Homology Modeling of Protein Structures

Homology models of proteins were built by using ModPipe (Pieper et al., 2014). The homology structure with the highest ModPipe quality score was selected for each protein according to the previously described criteria (Zhang et al., 2016; Liu et al., 2017).

### Structural Template for the Interaction Model

A total of 157,771 protein complexes involving 328,671 chains were collected from the Protein Data Bank (PDB) (Burley et al., 2017). The chain–chain binary interfaces of protein complexes were generated by PIBASE with an interatomic distance cutoff of 6.05 Å (Davis and Sali, 2005).

### Structure-Based Features

Structural alignment was used to find the closest PDB chains of homology models using TM-align with the cutoff score of 0.4 (Zhang and Skolnick, 2005). The interaction model of protein pair was created by superimposing the homology structures on their corresponding chains in the closest PDB template complex. Four structural features, including structural similarity (i.e., TM-score) and structural distance (i.e., root mean square deviation, RMSD) between protein homology models and their corresponding chains in the template, as well as the number and fraction of interacting residue pairs in the template that were preserved in the interaction model, were calculated for the prediction of PPIs. The detailed method for structural features refers to the previous study (Zhang et al., 2016).

### Prediction of Interolog and Domain-Based PPIs

The potential PPIs were predicted using interolog mapping. Each protein was blasted against the experimentally determined PPI datasets to identify homologs with an *E*-value of  $<10^{-5}$ , a sequence identity of  $>45\%$ , and an aligned sequence coverage of  $>50\%$ . The experimentally determined PPIs were derived from the BioGRID (Oughtred et al., 2021), IntAct (Orchard et al., 2014), MINT (Calderone et al., 2020), DIP (Salwinski et al., 2004), and BIND (Alfarano et al., 2005) databases (**Supplementary Table 1**).

The domains of each protein were identified by PfamScan against the Pfam database (Mistry et al., 2021). The interacting domains were identified based on the host–pathogen PPIs and collected from the 3 did (three-dimensional interacting domains) database (Mosca et al., 2014). When an interacting domain pair

were present in two proteins, the two proteins were expected to interact with each other.

## Performance Evaluation of Predictive Models

To evaluate model performance across different host–pathogen systems, we trained on the dataset in the training host–pathogen systems and evaluated the performance on the test host–pathogen system. In other words, the entire training dataset was partitioned into two parts, i.e., test set and training set. Test set contained dataset from one host and all its pathogens, and training set included all datasets related to the other remaining hosts and pathogens. We used the test set from seven hosts (i.e., *Mus musculus*, *Arabidopsis thaliana*, *Rattus norvegicus*, *Aedes aegypti*, *Bos taurus*, *Sus scrofa*, and *Gallus gallus*) and their pathogens to evaluate the model performance across host–pathogen systems, respectively. For each test host–pathogen system, the process was performed 10 times with the different negative training sets.

The 10-fold cross-validation method was used to evaluate the performance of different models. The training dataset was randomly divided into 10 subsets. Nine of them were combined to train the model, and the remaining one was used to test the model. The process was repeated 10 times with the different negative training sets, and the final result was the average performance of the 10 replicates. True positive rate (TPR) =  $TP/(TP + FN)$ , false positive rate (FPR) =  $FP/(FP + TN)$ , precision =  $TP/(TP + FP)$ , and F1 score =  $2 \times (\text{precision} \times \text{recall})/(\text{precision} + \text{recall})$  were used to evaluate the prediction performance.

## Prediction of PPIs Between Rice and Blast Fungus

We built the rice–blast fungus PPI classifier using the random forest from the scikit-learn library in Python. The interaction probability of each protein pair was computed using the optimized model. The protein pair with a probability greater than the threshold of 0.5 was considered to interact with each other.

## Identification of Avirulence Genes in Blast Fungal Genome

Gene sequences of avirulence effectors were extracted from the GenBank database (Meng et al., 2019). All predicted coding sequences of the blast fungal genome were then searched against the local avirulence sequence database to identify matches to the cloned genes. The parameters used for the sequence similarity search were  $\geq 95\%$  identity and 80% coverage of the avirulence effector genes.

## Analysis of Functional and Pathway Enrichment

Functional enrichment of rice genes was analyzed by using agriGO (Tian et al., 2017) by comparing the reference gene dataset of the rice genome with their False Discovery Rate (FDR) values. The pathway enrichment analysis of rice genes was performed using the Fisher's exact test implemented in a Perl

script against the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa et al., 2017).

## Network Analysis of Blast Resistance Genes

The main data on blast disease traits were collected from the China Rice Data Center ([http://www.ricedata.cn/gene/gene\\_pi.htm](http://www.ricedata.cn/gene/gene_pi.htm)). The blast resistance genes that were identified in the rice reference genome were used in the network analysis.

## Transcriptional Analysis of Rice Genes in Response to Blast Fungus

The RNA-seq datasets (accession: SRP079683) derived from rice with *M. oryzae* infection were downloaded from the NCBI SRA database (Bidzinski et al., 2016). The sequencing reads were processed by trimmomatic to remove the adapter sequences and low-quality reads (Bolger et al., 2014). The cleaned reads were mapped to the rice genome using hisat2 (Kim et al., 2019), and the aligned reads were counted with featureCounts (Liao et al., 2014). The differentially expressed genes were identified using DESeq2 with  $p < 0.01$  and at least 2-fold changes (Love et al., 2014).

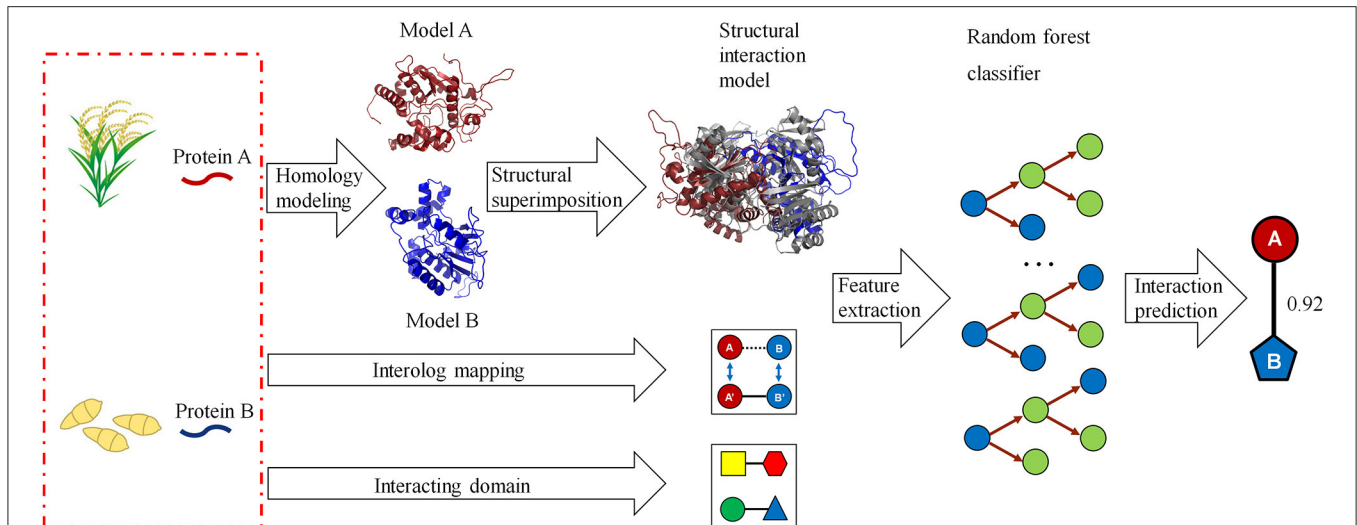
## RESULTS

### Inference of Host–Pathogen PPIs From Structure Information

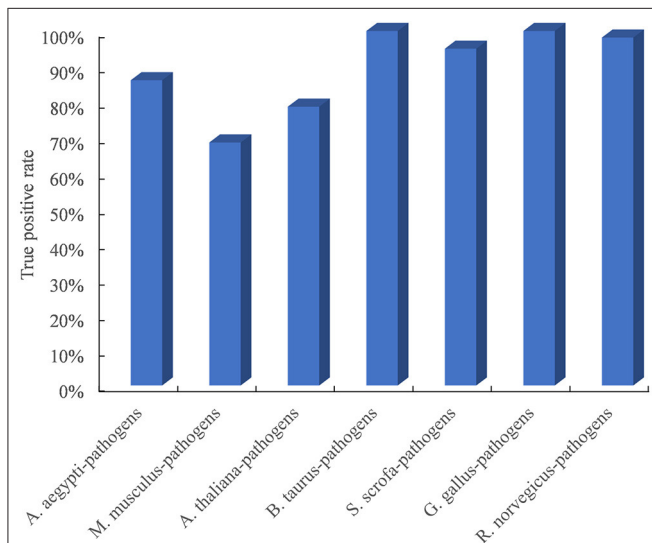
We developed a computational method for predicting host–pathogen PPIs based on structural information. The framework is illustrated in **Figure 1**. In brief, given a pair of proteins from host and pathogen, we first predicted protein structures using homology modeling and then searched for their closest PDB complex as a structural template. The interaction model for each protein pair was created by superimposing the homology structures on their corresponding chains in the template complex. Structural features, including structural similarity, structural distance, as well as the number and fraction of the conserved interacting residue pairs, were calculated from the interaction model. Finally, we combined structural evidences and sequence information to predict host–pathogen PPIs using the random forest-based classifiers.

### Structure-Based Method Accurately Discovers Plant–Pathogen PPIs

Due to the limited availability of known PPIs between plants and pathogens, cross-species performance is important for the model trained on the dataset in known host–pathogen systems to infer PPIs in a new host–pathogen system. We selected seven test host–pathogen systems to systematically evaluate model performance across host–pathogen systems. For each test, all datasets except from one selected host and all its pathogens were used to train the model, and the dataset from the selected host–pathogen system was used to evaluate the predictive model. As shown in **Figure 2**, the TPR of the method is higher than 68.6% for all test host–pathogen systems. The results indicated its robust



**FIGURE 1** | A computational framework for predicting protein–protein interactions (PPIs) between host and pathogen. Given a pair of potentially interacting proteins (A,B) from host and pathogen, three-dimensional structures for each protein pair were built by homology modeling and then searched for their closest PDB complex as a template by structural alignment. Four structure-based scores associated with the protein pair were calculated by superimposing the homology structures on their corresponding chains in the template complex. Finally, combining the structural evidences with the clues of homologous mapping and interacting domains to predict the interaction between proteins A and B using random forest-based classifiers.



**FIGURE 2** | Performance evaluation of predicting PPIs across host–pathogen systems. The entire training dataset was partitioned into a test set and a training set. The test set contained the dataset from one selected host and all its pathogens, and the training set included all datasets related to the other remaining hosts and pathogens. For each test host–pathogen system, the process was performed 10 times with different negative training sets.

cross-species prediction performance in discovering PPIs in new host–pathogen systems.

The interolog and domain-based methods are two widely used computational approaches for predicting host–pathogen PPIs. We thus evaluated the performance of our method with respect to these two methods on the test dataset of experimentally determined PPIs between *Arabidopsis* and all its pathogens.

**TABLE 1** | The performance of different approaches in detecting protein–protein interactions (PPIs) between *Arabidopsis* and pathogen.

Method	Accuracy (ACC)	True positive rate (TPR)	F1-score
Structure-based method	224/256 = 87.5%	101/128 = 78.9%	0.86
domain-based method	148/256 = 57.8%	20/128 = 15.6%	0.27
interolog-based method	151/256 = 59.0%	23/128 = 18.0%	0.30

The test dataset including 128 experimentally determined PPIs and 128 randomly shuffled protein pairs between *Arabidopsis* and pathogens.

Among the 128 *Arabidopsis*–pathogen PPIs with homology-modeled structures, 101 (78.9%) protein interactions could be successfully predicted by the structure-based method. However, only 23 and 20 *Arabidopsis*–pathogen PPIs were inferred by the homologous mapping and interacting domain pairs, respectively (Table 1). For the interaction between *Arabidopsis* and *Ralstonia solanacearum*, 20 of PPIs were detected by the structural similarity, while only 4 of them were supported with the interacting domains (Supplementary Table 2). These results indicated that the structure-based method outperformed the interolog and domain-based methods for identifying plant–pathogen PPIs.

## Proteome-Wide Prediction of PPIs Between Rice and Blast Fungus

We first trained the prediction model on a dataset consisting of positive and negative examples of equal size. Although the model worked relatively well with an accuracy of 94.13%, the



**TABLE 2** | Performance comparison of models trained with different ratios of positive and negative samples.

Positive:negative ratio	Accuracy (ACC) (%)	True positive rate (TPR) (%)	False positive rate (FPR) (%)
1:1 (10148: 10148)	94.13	93.63	5.37
1:5 (10148: 50740)	93.09	84.14	5.12
1:10 (10148: 101480)	94.17	66.30	3.04
1:20 (10148: 202960)	96.16	42.79	1.17
1:30 (10148: 304440)	97.30	32.60	0.54
1:50 (10148: 507400)	98.32	23.84	0.19
1:60 (10148: 608880)	98.58	21.75	0.14
<b>1:70 (10148: 710360)</b>	<b>98.78</b>	<b>20.77</b>	<b>0.10</b>
1:80 (10148: 811840)	98.93	19.66	0.08
1:90 (10148: 913320)	99.04	18.90	0.07
1:100 (10148: 1014800)	99.13	18.30	0.06

The bold values mean the optimized positive to negative ratio of training set with a relatively high TPR and a low level of FPR.

FPR of 5.37% would result in a large number of false positive interactions in the proteome-wide prediction of PPIs between rice and blast fungus. Thus, we reduced the FPR by expanding the size of the negative examples in the training dataset. When the ratio of positive to negative samples was adjusted to 1:70 in the training dataset, the FPR was decreased to the expected level of 0.10%, while the TPR remained at the relatively high level of 20.77% (Table 2).

To fill the gap between the number of protein sequences and 3D structures, protein structures for rice and blast fungus proteomes were predicted by using homology modeling. These predicted structures contained 32,170 and 2,910 models, covering 82.8 and 83.3% of rice and blast fungus secreted/transmembrane proteomes, respectively. Interaction models for 21,021,571 rice–blast fungus protein pairs were then created by superimposing the homology structures on their corresponding PDB templates.

The optimized host–pathogen prediction model was used to scan all rice–blast fungus protein pairs with the interaction models, resulting in a total of 2,018 PPIs between 1,344 rice proteins and 418 blast fungus proteins (Figure 3A, Supplementary Table 3). Of the predicted PPIs, only 29.9% (604) of interactions were supported with the evidences of sequence similarity and interacting domain pairs. Moreover, we found that 17 predicted PPIs could be identified by the previous study with the sequence-based method (Ma et al., 2019). These results indicated that the structure-based method could efficiently discover new rice–blast fungus PPIs beyond those interactions inferred from sequence similarity.

As expected, the rice–blast fungus PPI network exhibited scale-free properties similar to those of other biological networks (Supplementary Figure 1). It was interesting that the blast fungus proteins had more connections than rice proteins in the PPI network. One blast fungus protein averagely had five interacting partners from rice, while one rice protein interacted with around two blast fungus proteins (Figure 3B). Approximately, 12% (50) of blast fungus proteins had at least 10 rice interactors, and the pathogen protein with the highest degree was predicted to interact with 143 rice partners in the network.

The result meant that the potential pathogen-associated proteins had a higher degree than the resistance-associated proteins in the rice–blast fungus PPI network.

## Network Captures Key Components in Rice–Blast Fungus Pathosystem

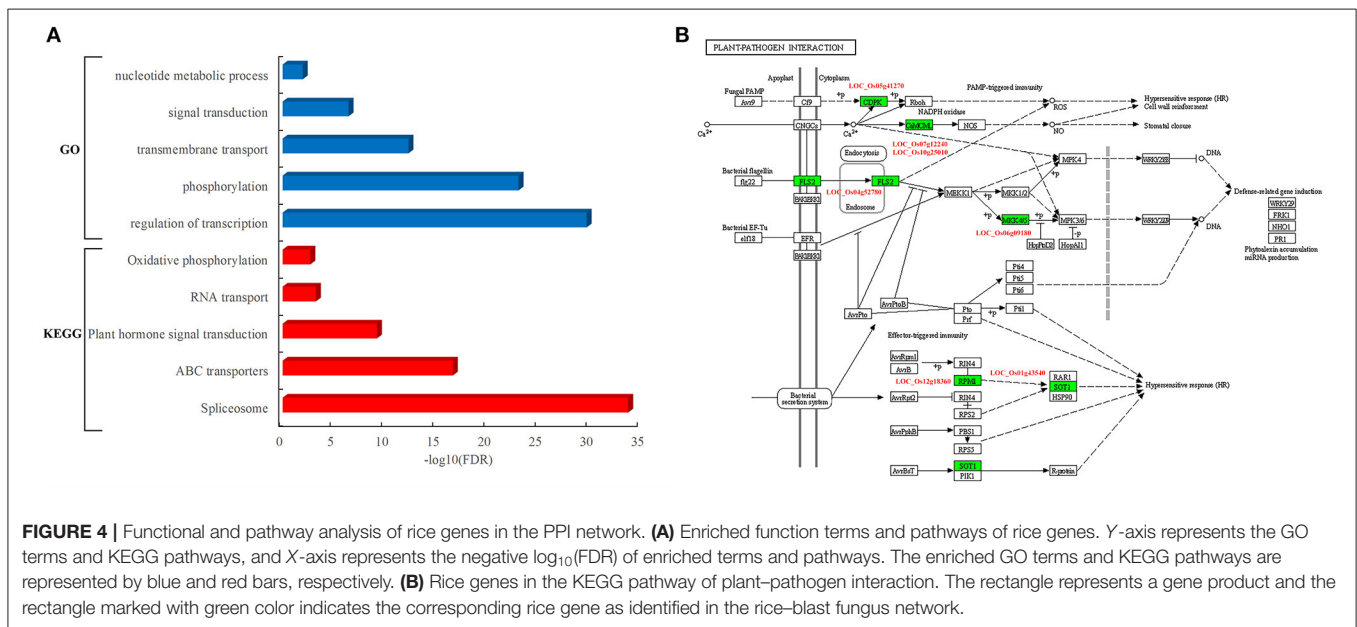
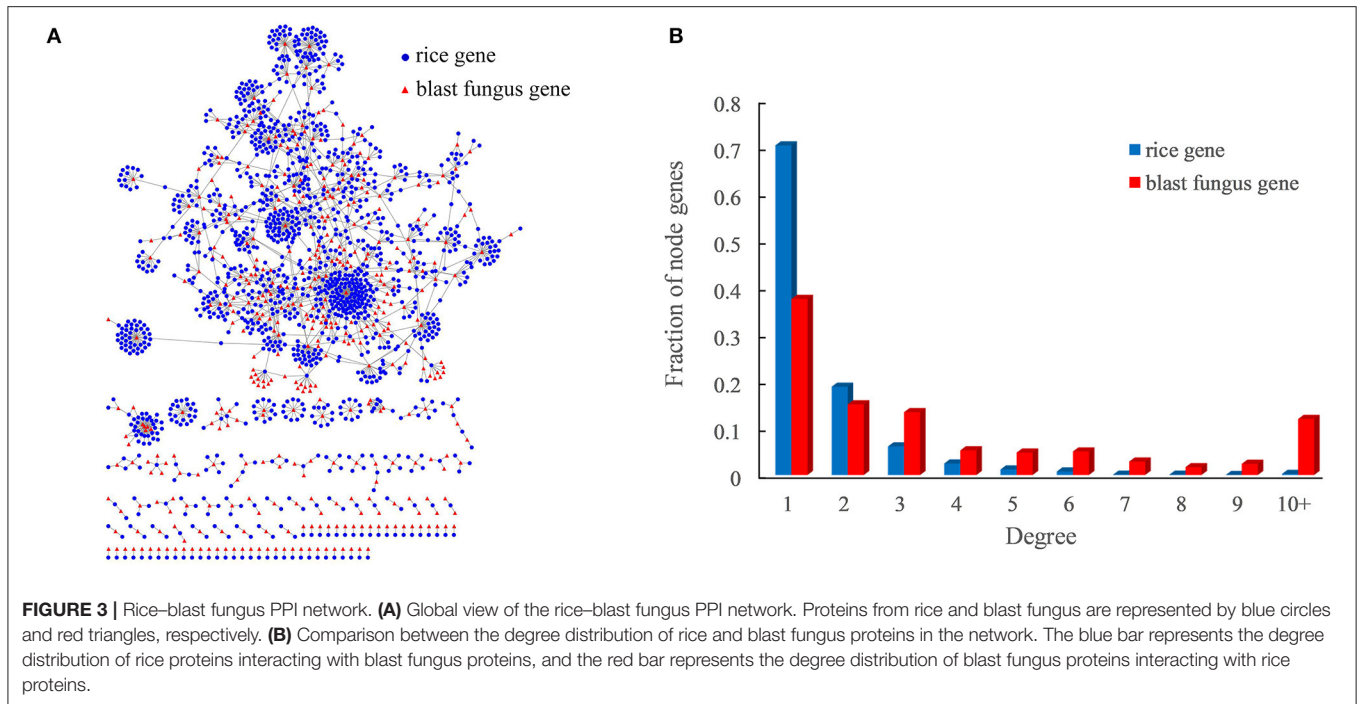
The major components involved in rice–blast fungus interactions include resistance genes from rice and avirulence effectors from *M. oryzae*. Thus, we examined whether the PPI network could predict these key components in rice–blast fungus pathosystem. Currently, about two dozens of blast resistance genes have been cloned and characterized in rice. Among the cloned resistance genes, 12 genes were distributed in the rice reference genome of *O. sativa* spp. *japonica* cv. Nipponbare (Supplementary Table 4). Two resistance genes, namely, pi-d2 (*LOC\_Os06g29810*) and pi-ta (*LOC\_Os12g18360*), were successfully predicted in the PPI network, ~7-fold enrichment in comparison with that of the whole genome. In addition, a total of 13 avirulence effector genes have been cloned from *M. oryzae*, six of which have the corresponding matches in the reference genome sequence (Supplementary Table 5). One avirulence effector, AVR-Pik (MGG\_15972), was detected in the PPI network. The AVR-Pik effector was predicted to interact with four resistance-associated proteins (i.e., *LOC\_Os02g37290*, *LOC\_Os02g37300*, *LOC\_Os02g37320*, and *LOC\_Os04g39380*) containing a heavy metal-associated domain, which have been validated by experimental measurements (De la Concepcion et al., 2018, 2019). These results indicated the powerful performance of the PPI network in capturing major components in rice–blast fungus interactions.

## Functional Analysis of Rice Protein Involved in the Interaction

To determine the function of rice proteins interacted with pathogen proteins, the GO analysis of these proteins in the network was carried out. These resistance-associated proteins were preferentially involved in specific biological processes such as transcriptional regulation, phosphorylation, transmembrane transport, signal transduction, and nucleotide metabolic process (Figure 4A). Furthermore, the pathway analysis showed that these rice proteins are significantly enriched in the pathways of spliceosome, ATP-binding cassette (ABC) transporters, plant hormone signal transduction, RNA transport, and oxidative phosphorylation (Figure 4A). It was worth to notice that seven rice proteins acted as core components in the pathway of plant–pathogen interactions, including *LOC\_Os04g52780*, known as FLS2, an immune receptor, which could activate plant immune response by recognizing flagellin proteins of bacterial pathogens (Figure 4B).

## Network-Based Prediction of Blast Resistance Gene

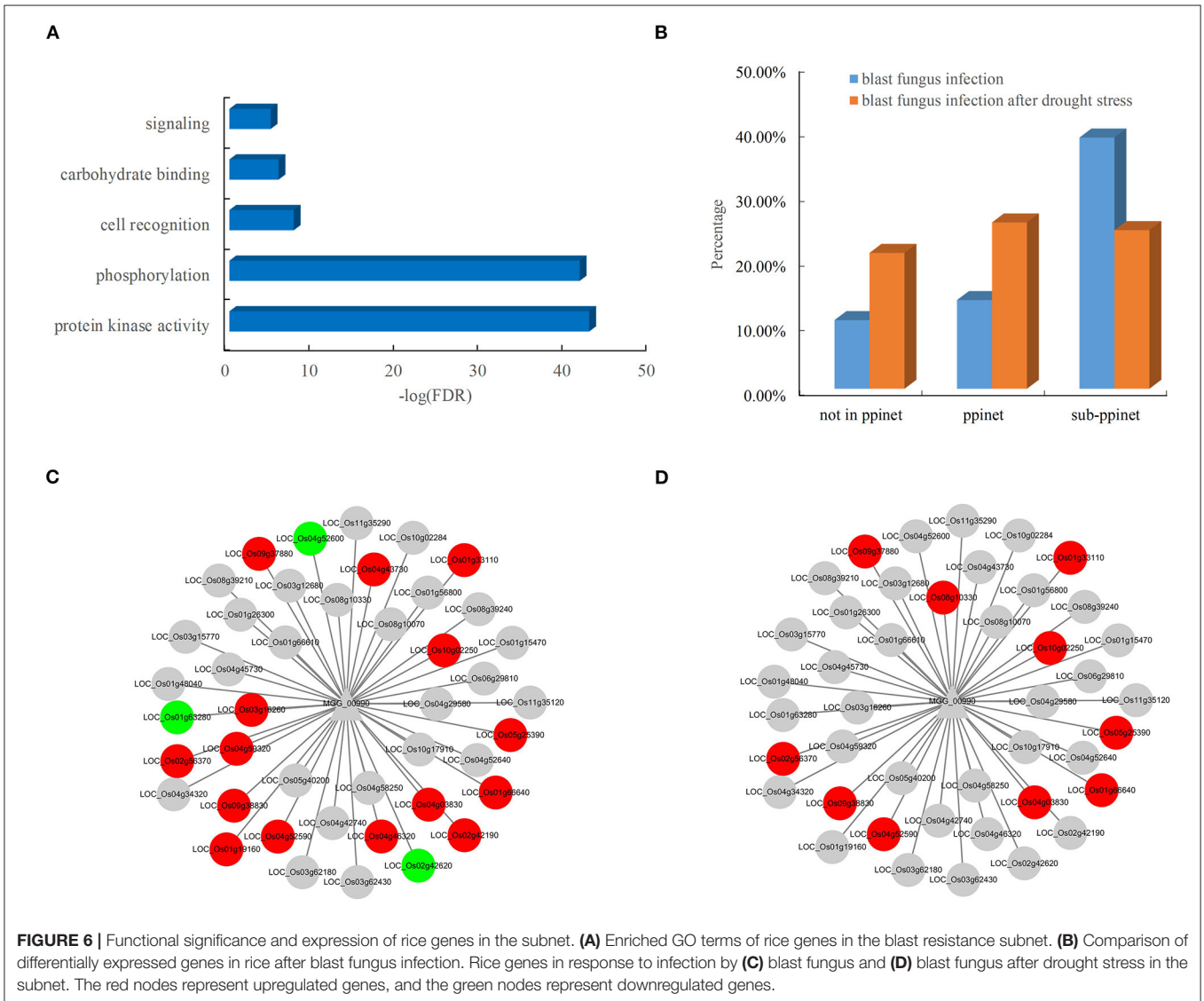
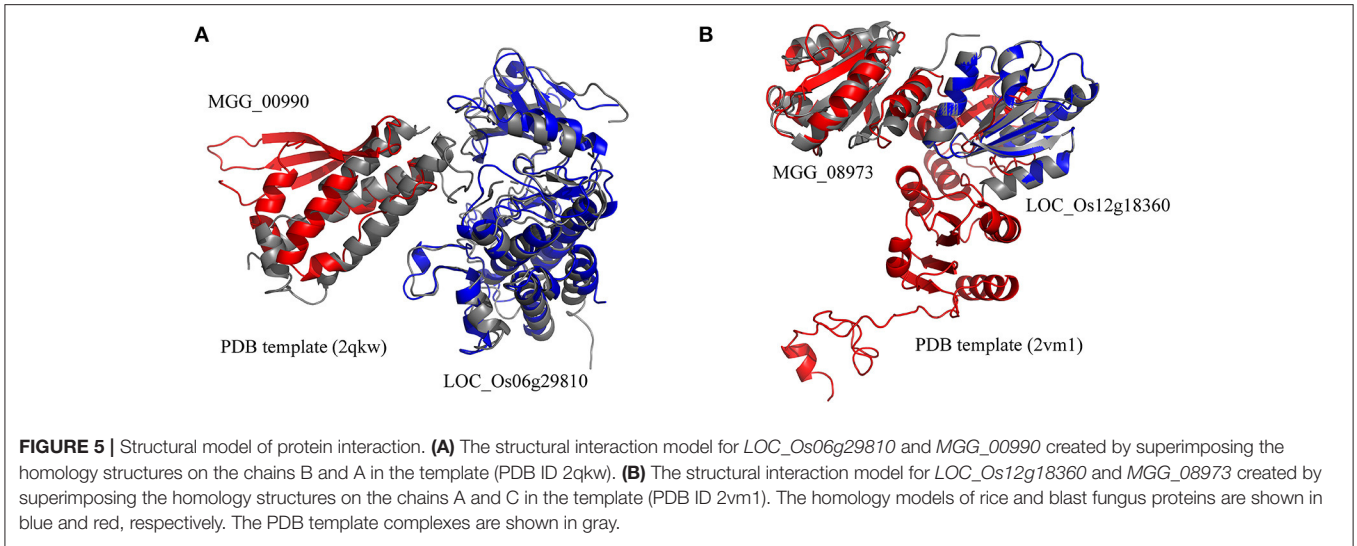
The rice blast resistance gene *LOC\_Os06g29810* was predicted to interact with pathogen protein *MGG\_00990* based on the structural interaction model created by superimposing the homology structures on the template of the AvrPto–Pto complex (Figure 5A). The homology model of *MGG\_00990* was structurally similar to the pathogen effector AvrPto



in the template complex, while the homology model of *LOC\_Os06g29810* was structurally close to the plant resistance protein Pto (Xing et al., 2007). The another rice blast resistance gene *LOC\_Os12g18360* could interact with the pathogen gene *MGG\_08973*, which was inferred from the structural similarity of the homology models to the chains in the structural template of thioredoxin in barley (Figure 5B) (Maeda et al., 2008).

The availability of the PPI network allowed the systematic discovery of novel blast resistance genes using the guilt-by-association method. In addition to the two blast resistance genes, 47 rice genes were found to interact with the two

pathogen-associated genes, *MGG\_00990* and *MGG\_08973*, in the PPI network. The functional analysis revealed that these blast resistance candidates mostly encoded receptor-like cytoplasmic kinases involved in the biological processes of phosphorylation and signaling in rice (Figure 6A). Furthermore, the analysis of gene expression showed that these candidates were preferentially responsive to the infection of blast fungus (Figure 6B). Among the interacting partners of *MGG\_00990*, 18 resistance-associated genes were differentially expressed after blast fungus infection (Figure 6C), while 10 genes were significantly induced by the infection blast fungus after drought stress (Figure 6D). These



results indicated the good performance of the PPI network in discovering rice genes associated with blast resistance.

## DISCUSSION

Determining PPIs is an essential step toward understanding the underlying biological mechanism against pathogen infection in plants (Muthamilarasan and Prasad, 2013). In this study, we presented a computational method for predicting plant-pathogen PPIs based on structural information. Using seven host-pathogen systems, we demonstrated that the structure-based method was powerful in discovering protein interactions across different host-pathogen systems. This advantage was of vital importance for machine learning-based method for predicting PPIs in new plant-pathogen systems based on prior knowledge obtained from known interacting proteins in other host-pathogen systems.

Although many experimentally validated host-pathogen PPIs have been deposited in the database, most of these PPIs (42,972 out of 45,200) focus on the protein interactions between humans and pathogens (Ammari et al., 2016). Due to the limited availability of plant-pathogen PPIs, many plant-specific pathogen effector proteins usually failed to identify any homolog in the known inter-species interactions using homologous sequence mapping (Yang et al., 2019). In addition, the intra-species PPIs from model organisms have been usually used to infer plant-pathogen PPIs (Li et al., 2012; Sahu et al., 2014). However, evolutionary differences between inter-species and intra-species PPI interfaces would limit the performance of plant-pathogen PPIs by transferring interactions across species (Franzosa and Xia, 2011). These weaknesses could be partially overcome by utilizing structural information. The developed method uses the structural similarity between proteins as a bridge to identify new interactions across plant-pathogen systems. Compared with the sequence-based methods, this structure-based approach enabled us to discover new interactions between plant and pathogen proteins that lacked significant sequence similarity with a known interaction template.

*Magnaporthe oryzae* is a notorious plant pathogen that causes the most destructive diseases of rice in the world. The prediction and analysis of PPIs are valuable in deciphering the molecular mechanisms of rice-blast fungus interactions. In this study, we generated a global rice-blast fungus PPI network that consisted of 2,018 interacting protein pairs involving 1,344 rice proteins and 418 blast fungus proteins. Over 70% of PPIs between rice and blast fungus were inferred from structural similarity, which greatly expanded the landscape of the rice-blast fungus PPI network. Compared with the previous PPI network, 17 of PPIs were identified by the structure-based method (Ma et al., 2019). Although the number of common PPIs was relatively small, the results were significantly overlapped between the two independent studies (i.e., Fisher's exact test  $p < 5.6e-39$ ). Moreover, we noted that blast fungus proteins had more interacting partners than rice proteins in the network. Our findings were coherent with the sequence-based studies in which

a few pathogen-associated proteins were involved in the plant-pathogen interactions (Li et al., 2012; Sahu et al., 2014; Ma et al., 2019). This is likely to be the result of the coevolutionary arms races, in which pathogens mutate genes extensively to infect their hosts, while plants defend against pathogen attacks by expanding gene families (Stahl and Bishop, 2000; Dangl and McDowell, 2006).

Despite the advances made in molecular mechanisms of rice resistance to blast fungus, many aspects of the rice immunity system remain obscure (Li et al., 2019). The rice-blast fungus PPI network showed that the AVR-Pik effector was successfully predicted to interact with four rice proteins, which have been validated by experimental approaches (De la Concepcion et al., 2018, 2019). In addition to the avirulence effector, two rice blast resistance genes were also identified in the network. Using the guilt-by-association method, we identified 47 candidate blast resistance genes in the PPI network. The majority of these genes that encoded receptor-like cytoplasmic kinases were involved in the response to the infection of blast fungus (Bidzinski et al., 2016). The PPI network provided a global map to help accelerate the identification of blast resistance genes and advance our understanding of the molecular mechanisms of plant-pathogen interactions.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

## AUTHOR CONTRIBUTIONS

CZ, LZhao, and LZhang designed the project and carried out the model training and computational validation. YL performed bioinformatic analyses. FS made substantial contributions to data collection. CZ, LZhao, and LZhang wrote the manuscript. All authors approved the manuscript.

## FUNDING

This work was supported by the National Natural Science Foundation of China (Grant No. 31771343). The data computing was supported by Center for HPC, Shanghai Jiao Tong University.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2021.690124/full#supplementary-material>

**Supplementary Figure 1** | Degree distribution of the node proteins.

**Supplementary Table 1** | Availability of protein-protein interactions (PPIs) in model organisms from various databases.

**Supplementary Table 2** | The experimentally determined PPIs between *Arabidopsis* and pathogens.

**Supplementary Table 3** | The predicted PPIs between rice and blast fungus.



**Supplementary Table 4** | The isolated dominant blast resistance genes in rice reference genome.

**Supplementary Table 5** | The isolated avirulence genes in *Magnaporthe oryzae* genome.

## REFERENCES

- Alfarano, C., Andrade, C. E., Anthony, K., Bahroos, N., Bajec, M., Bantoft, K., et al. (2005). The biomolecular interaction network database and related tools 2005 update. *Nucleic Acids Res.* 33, D418–D424. doi: 10.1093/nar/gki051
- Almagro Armenteros, J. J., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., et al. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* 37, 420–423. doi: 10.1038/s41587-019-0036-z
- Ammari, M. G., Gresham, C. R., McCarthy, F. M., and Nanduri, B. (2016). HPIDB 2.0: a curated database for host-pathogen interactions. *Database* 2016:baw103. doi: 10.1093/database/baw103
- Bidzinski, P., Ballini, E., Ducasse, A., Michel, C., Zuluaga, P., Genga, A., et al. (2016). Transcriptional basis of drought-induced susceptibility to the rice blast fungus *Magnaporthe oryzae*. *Front. Plant Sci.* 7:1558. doi: 10.3389/fpls.2016.01558
- Bolger, A. M., Lohse, M., Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 30:2114–2120. doi: 10.1093/bioinformatics/btu170
- Burley, S. K., Berman, H. M., Kleywegt, G. J., Markley, J. L., Nakamura, H., and Velankar, S. (2017). Protein Data Bank (PDB): the single global macromolecular structure archive. *Methods Mol. Biol.* 1607, 627–641. doi: 10.1007/978-1-4939-7000-1\_26
- Calderone, A., Iannuccelli, M., Peluso, D., and Licata, L. (2020). Using the MINT database to search protein interactions. *Curr. Protoc. Bioinformatics.* 69:e93. doi: 10.1002/cpbi.93
- Cao, F. Y., Khan, M., Taniguchi, M., Mirmiran, A., Moeder, W., Lumba, S., et al. (2019). A host-pathogen interactome uncovers phytopathogenic strategies to manipulate plant ABA responses. *Plant J.* 100, 187–198. doi: 10.1111/tpj.14425
- Cesari, S., Thilliez, G., Ribot, C., Chalvon, V., Michel, C., Jauneau, A., et al. (2013). The rice resistance protein pair RGA4/RGA5 recognizes the *Magnaporthe oryzae* effectors AVR-Pia and AVR1-CO39 by direct binding. *Plant Cell.* 25, 1463–1481. doi: 10.1105/tpc.112.107201
- Dangl, J. L., and McDowell, J. M. (2006). Two modes of pathogen recognition by plants. *Proc. Natl. Acad. Sci. U.S.A.* 103:8575–8576. doi: 10.1073/pnas.0603183103
- Davis, F. P., and Sali, A. (2005). PIBASE: a comprehensive database of structurally defined protein interfaces. *Bioinformatics* 21, 1901–1907. doi: 10.1093/bioinformatics/bti277
- De la Concepcion, J. C., Franceschetti, M., MacLean, D., Terauchi, R., Kamoun, S., Banfield, M. J. (2019). Protein engineering expands the effector recognition profile of a rice NLR immune receptor. *Elife* 8:e47713. doi: 10.7554/eLife.47713
- De la Concepcion, J. C., Franceschetti, M., Maqbool, A., Saitoh, H., Terauchi, R., Kamoun, S., et al. (2018). Polymorphic residues in rice NLRs expand binding and response to effectors of the blast pathogen. *Nat. Plants* 4, 576–585. doi: 10.1038/s41477-018-0194-x
- Dean, R. A., Talbot, N. J., Ebbole, D. J., Farman, M. L., Mitchell, T. K., Orbach, M. J., et al. (2005). The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature* 434, 980–986. doi: 10.1038/nature03449
- Deng, M., Mehta, S., Sun, F., and Chen, T. (2002). Inferring domain-domain interactions from protein-protein interactions. *Genome Res.* 12, 1540–1548. doi: 10.1101/gr.153002
- Franzosa, E. A., and Xia, Y. (2011). Structural principles within the human-virus protein-protein interaction network. *Proc. Natl. Acad. Sci. U.S.A.* 108, 10538–10543. doi: 10.1073/pnas.1101440108
- Horton, P., Park, K. J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C. J., et al. (2007). WoLF PSORT: protein localization predictor. *Nucleic Acids Res.* 35, W585–W587. doi: 10.1093/nar/gkm259
- International Rice Genome Sequencing Project. (2005). The map-based sequence of the rice genome. *Nature* 436, 793–800. doi: 10.1038/nature03895
- Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 45, D353–D361. doi: 10.1093/nar/gkw1092
- Kanzaki, H., Yoshida, K., Saitoh, H., Fujisaki, K., Hirabuchi, A., Alaux, L., et al. (2012). Arms race co-evolution of *Magnaporthe oryzae* AVR-Pik and rice Pik genes driven by their physical interactions. *Plant J.* 72, 894–907. doi: 10.1111/j.1365-313X.2012.05110.x
- Kim, D., Paggi, J. M., Park, C., Bennett, C., and Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* 37, 907–915. doi: 10.1038/s41587-019-0201-4
- Li, W., Chern, M., Yin, J., Wang, J., and Chen, X. (2019). Recent advances in broad-spectrum resistance to the rice blast disease. *Curr. Opin. Plant Biol.* 50, 114–120. doi: 10.1016/j.pbi.2019.03.015
- Li, Z. G., He, F., Zhang, Z., and Peng, Y. L. (2012). Prediction of protein-protein interactions between *Ralstonia solanacearum* and *Arabidopsis thaliana*. *Amino Acids.* 42, 2363–2371. doi: 10.1007/s00726-011-0978-z
- Liao, Y., Smyth, G. K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930. doi: 10.1093/bioinformatics/btt656
- Liu, S., Liu, Y., Zhao, J., Cai, S., Qian, H., Zuo, K., et al. (2017). A computational interactome for prioritizing genes associated with complex agronomic traits in rice (*Oryza sativa*). *Plant J.* 90, 177–188. doi: 10.1111/tpj.13475
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550. doi: 10.1186/s13059-014-0550-8
- Ma, S., Song, Q., Tao, H., Harrison, A., Wang, S., Liu, W., et al. (2019). Prediction of protein-protein interactions between fungus (*Magnaporthe grisea*) and rice (*Oryza sativa* L.). *Brief Bioinform.* 20, 448–456. doi: 10.1093/bib/bbx132
- Maeda, K., Häggglund, P., Finnie, C., Svensson, B., and Henriksen, A. (2008). Crystal structures of barley thioredoxin h isoforms HvTrxh1 and HvTrxh2 reveal features involved in protein recognition and possibly in discriminating the isoform specificity. *Protein Sci.* 17, 1015–1024. doi: 10.1110/ps.083460308
- Marcotte, E. M., Pellegrini, M., Ng, H. L., Rice, D. W., Yeates, T. O., and Eisenberg, D. (1999). Detecting protein function and protein-protein interactions from genome sequences. *Science* 285, 751–753. doi: 10.1126/science.285.5428.751
- Mariano, R., and Wuchty, S. (2017). Structure-based prediction of host-pathogen protein interactions. *Curr. Opin. Struct. Biol.* 44, 119–124. doi: 10.1016/j.sbi.2017.02.007
- Matthews, L. R., Vaglio, P., Reboul, J., Ge, H., Davis, B. P., Garrels, J., et al. (2001). Identification of potential interaction networks using sequence-based searches for conserved protein-protein interactions or “interologs”. *Genome Res.* 11, 2120–2126. doi: 10.1101/gr.205301
- Meng, Q., Gupta, R., Min, C. W., Kwon, S. W., Wang, Y., Je, B. I., et al. (2019). Proteomics of rice-*Magnaporthe oryzae* interaction: what have we learned so far? *Front. Plant Sci.* 10:1383. doi: 10.3389/fpls.2019.01383
- Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., et al. (2021). Pfam: the protein families database in 2021. *Nucleic Acids Res.* 49, D412–D419. doi: 10.1093/nar/gkaa913
- Möller, S., Croning, M. D., and Apweiler, R. (2001). Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics* 17, 646–653. doi: 10.1093/bioinformatics/17.7.646
- Mosca, R., Céol, A., Stein, A., Olivella, R., and Aloy, P. (2014). 3did: a catalog of domain-based interactions of known three-dimensional structure. *Nucleic Acids Res.* 42, D374–D379. doi: 10.1093/nar/gkt887
- Mukhtar, M. S., Carvunis, A. R., Dreze, M., Epple, P., Steinbrenner, J., Moore, J., et al. (2011). Independently evolved virulence effectors converge onto hubs in a plant immune system network. *Science* 333, 596–601. doi: 10.1126/science.1203659
- Muthamilarasan, M., and Prasad, M. (2013). Plant innate immunity: an updated insight into defense mechanism. *J. Biosci.* 38, 433–449. doi: 10.1007/s12038-013-9302-2

- Orchard, S., Ammari, M., Aranda, B., Breuza, L., Briganti, L., Broackes-Carter, F., et al. (2014). The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res.* 42, D358–D363. doi: 10.1093/nar/gkt1115
- Oughtred, R., Rust, J., Chang, C., Breitkreutz, B. J., Stark, C., Willems, A., et al. (2021). The BioGRID database: a comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Sci.* 30, 187–200. doi: 10.1002/pro.3978
- Pellegrini, M., Marcotte, E. M., Thompson, M. J., Eisenberg, D., and Yeates, T. O. (1999). Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc. Natl. Acad. Sci. U.S.A.* 96, 4285–4288. doi: 10.1073/pnas.96.8.4285
- Pieper, U., Webb, B. M., Dong, G. Q., Schneidman-Duhovny, D., Fan, H., Kim, S. J., et al. (2014). ModBase, a database of annotated comparative protein structure models and associated resources. *Nucleic Acids Res.* 42, D336–D346. doi: 10.1093/nar/gkt1144
- Sahu, S. S., Weirick, T., and Kaundal, R. (2014). Predicting genome-scale Arabidopsis-Pseudomonas syringae interactome using domain and interolog-based approaches. *BMC Bioinform.* 11:S13. doi: 10.1186/1471-2105-15-S11-S13
- Salwinski, L., Miller, C. S., Smith, A. J., Pettit, F. K., Bowie, J. U., and Eisenberg, D. (2004). The database of interacting proteins: 2004 update. *Nucleic Acids Res.* 32, D449–D451. doi: 10.1093/nar/gkh086
- Stahl, E. A., and Bishop, J. G. (2000). Plant-pathogen arms races at the molecular level. *Curr. Opin. Plant Biol.* 3, 299–304. doi: 10.1016/S1369-5266(00)00083-2
- Tanwar, H., and George Priya Doss, C. (2018). Computational resources for predicting protein-protein interactions. *Adv. Protein Chem. Struct. Biol.* 110, 251–275. doi: 10.1016/bs.apcsb.2017.07.006
- Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., et al. (2017). agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res.* 45, W122–W129. doi: 10.1093/nar/gkx382
- Weßling, R., Eppele, P., Altmann, S., He, Y., Yang, L., Henz, S. R., et al. (2014). Convergent targeting of a common host protein-network by pathogen effectors from three kingdoms of life. *Cell Host Microbe.* 16, 364–375. doi: 10.1016/j.chom.2014.08.004
- Xing, W., Zou, Y., Liu, Q., Liu, J., Luo, X., Huang, Q., et al. (2007). The structural basis for activation of plant immunity by bacterial effector protein AvrPto. *Nature* 449, 243–247. doi: 10.1038/nature06109
- Yang, S., Li, H., He, H., Zhou, Y., and Zhang, Z. (2019). Critical assessment and performance improvement of plant-pathogen protein-protein interaction prediction methods. *Brief Bioinform.* 20, 274–287. doi: 10.1093/bib/bbx123
- Zhang, F., Liu, S., Li, L., Zuo, K., Zhao, L., and Zhang, L. (2016). Genome-wide inference of protein-protein interaction networks identifies crosstalk in abscisic acid signaling. *Plant Physiol.* 171, 1511–1522. doi: 10.1104/pp.16.00057
- Zhang, Q. C., Petrey, D., Deng, L., Qiang, L., Shi, Y., Thu, C. A., et al. (2012). Structure-based prediction of protein-protein interactions on a genome-wide scale. *Nature* 490, 556–560. doi: 10.1038/nature11503
- Zhang, Y., and Skolnick, J. (2005). TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* 33, 2302–2309. doi: 10.1093/nar/gki524
- Zhao, J., Lei, Y., Hong, J., Zheng, C., and Zhang, L. (2019). AraPPINet: An Updated Interactome for the Analysis of Hormone Signaling Crosstalk in *Arabidopsis thaliana*. *Front. Plant Sci.* 10:870. doi: 10.3389/fpls.2019.00870

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Zheng, Liu, Sun, Zhao and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.