

FPPI: *Fusarium graminearum* Protein–Protein Interaction Database

Xing-Ming Zhao,^{*,†,‡} Xiao-Wei Zhang,[†] Wei-Hua Tang,[†] and Luonan Chen^{‡,¶}

Institute of Plant Physiology and Ecology, Chinese Academy of Sciences, Shanghai 200032, China, Institute of Systems Biology, Shanghai University, Shanghai 200444, China, and Department of Electrical Engineering and Electronics, Osaka Sangyo University, Osaka 574-8530, Japan

Received May 09, 2009

The fungal pathogen *Fusarium graminearum* (telomorph *Gibberella zeae*) is the causal agent of several destructive crop diseases. Identifying interactions among *F. graminearum* proteins and understanding their functions can provide insights into pathogenic mechanisms underlying *F. graminearum*–host interactions. *F. graminearum* protein–protein interaction (FPPI) database provides comprehensive information of protein–protein interactions (PPIs) of *F. graminearum* predicted based on both interologs from several PPI databases of seven species and domain–domain interactions experimentally determined based on protein structures. FPPI contains 223 166 interactions among 7406 proteins for *F. graminearum*. To the best of our knowledge, it is the first PPI map for this destructive fungus, which is thereby expected to shed light on biological functions of *F. graminearum* proteins. The predicted interactome covers about 52% of the whole *F. graminearum* proteome, and each interaction is assigned a score as the confidence for the predicted PPI. In particular, we constructed a core PPI data set with high confidence that consists of 27 102 interactions and 3745 proteins. To verify the reliability of the predicted interactome, we conducted yeast two-hybrid experiments over 3 randomly selected predictions from the core PPI data set, among which one pair of proteins was confirmed to indeed interact with each other, thereby proving high confidence on the core PPI data set. In addition, FPPI contains other functional information for *F. graminearum* genes, including homologues in other species deposited in different databases and the inferred functional characteristics, and so on. We further constructed an intuitive query interface for the database that provides easy access to the important features of proteins. In summary, FPPI is a rich source of information for system-level understanding of gene functions and biological processes in *F. graminearum*. Public access to the FPPI database is available at <http://csb.shu.edu.cn/fppi>.

Keywords: Domain–domain interaction • *Fusarium graminearum* • interolog • molecular pathogenicity • protein–protein interaction • wheat headblight

Introduction

Fusarium graminearum (*Gibberella zeae*) is the most common causal agent of *Fusarium* head blight of wheat and barley, and *Gibberella* stalk rot of maize.¹ It is estimated that *F. graminearum* causes economical losses of \$3 billion in the U.S. between 1991 and 1996.² Therefore, it is necessary to investigate the mechanism underlying the pathogenic process of this fungus, in order to facilitate the searching for an efficient way to control it. Although some pathogenicity genes have been identified for this fungus, such as pathogenicity genes deposited in PHI-base,³ the molecular mechanism by which *F. graminearum* overcomes plant defense barriers and causes a disease is still largely unknown. On the basis of our current understanding of pathogenicity of model pathogens,⁴ *F.*

graminearum is thought to organize a complex network of proteins and other molecules, including those that might be secreted into host cells, to adapt the life inside its host plant. Therefore, revealing the functions of proteins can provide insights into biological processes of *F. graminearum*, and help to understand the pathogenic process of this destructive fungus. Although the whole genome of *F. graminearum* has been sequenced and partly annotated,⁵ there are only about 4000 proteins that have been annotated in MIPS *F. graminearum* Genome Database (FGDB),⁶ which is far from complete. On the other hand, proteins generally exert their functions by interacting with each other in living organisms. Therefore, protein–protein interactions (PPIs) not only provide the basic building-blocks to construct protein interaction networks, but also give important hints about protein functions.

Recently, a large amount of protein–protein interaction data have been accumulated through high-throughput technologies. For example, there are interactome data for *Saccharomyces cerevisiae*,⁷ *Caenorhabditis elegans*,⁸ *Drosophila melanogaster*,^{9,10} *Homo sapiens*,¹¹ and so on. These interactomes are becoming

* To whom correspondence should be addressed. 99 Shangda Road, Institute of Systems Biology, Shanghai University, Shanghai 200444, China. E-mail: xm_zhao@shu.edu.cn.

[†] Institute of Plant Physiology and Ecology, CAS.

[‡] Shanghai University.

[¶] Osaka Sangyo University.

invaluable resources in systems biology by analyzing biomolecular networks formed by pairwise protein–protein interactions,¹² which provide biological insights into signaling and regulatory pathways as well as cellular functions of proteins. Unfortunately, there is no public accessible interactome for *F. graminearum* now, although it is strongly demanded. Detecting protein–protein interactions through high-throughput approach (e.g., yeast two-hybrid) is not feasible for this fungus due to the cost and time-consuming experiments. Therefore, computational methods provide an alternative way to predict protein–protein interactions for this fungus, and the preliminary computation results can serve as guideline for future experiment verification. In literature, a number of computational methods have been developed to predict PPIs based on various attributes and data types, such as interaction-ortholog (interolog),¹³ gene expression profiles,¹⁴ gene annotations,¹⁵ domain compositions,^{16–18} and coevolution.¹⁹ Among all computational methods, the interolog method is widely used, and it deduces interactions from experimentally determined interactomes of other species. In interolog approach, a pair of proteins in the organism of interest is regarded as interacting pair if their corresponding orthologs in other species interact with each other accordingly. The interolog method depends on accurately predicted ortholog genes, and the predicted interactome is limited to interactions among the most conserved proteins. On the other hand, PPIs are generally assumed to be mediated by domain–domain interactions (DDIs).¹⁶ With experimentally determined DDIs, we can predict PPIs for proteins that contain interacting domains (even without interacting orthologs), thereby complementing the interolog approach.

In this work, we predicted 223 166 interactions among 7406 proteins for *F. graminearum* based on interologs and DDIs. Although the predicted PPIs cover only about 52% of the whole *F. graminearum* proteome, it is the first PPI map for this destructive fungus, thereby shedding light on biological functions of *F. graminearum* proteins. Furthermore, we developed a score for each interaction as the confidence to evaluate the predicted PPIs. In particular, we constructed a core PPI data set with high confidence that consists of 27 102 interactions and 3745 proteins. By conducting yeast-two hybrid experiments, we verified one predicted interaction from 3 randomly selected predictions in the core data set, thereby confirming the reliability of the FPPI database. In addition, FPPI contains other functional information for *F. graminearum* genes, including homologues in other species deposited in different databases and the inferred functional characteristics, and so on. We further constructed an intuitive query interface that provides easy access to the important features of *F. graminearum* proteins. In summary, FPPI is a rich source of information for system-level understanding of gene functions and biological processes in *F. graminearum*. Public access to the FPPI database is available at <http://csb.shu.edu.cn/fppi>.

Methods

In this section, we present two approaches for predicting protein–protein interactions for *F. graminearum* based on interologs and domain–domain interactions, respectively. Figure 1 shows the schematic overview of predicting *F. graminearum* protein–protein interactions based on interologs

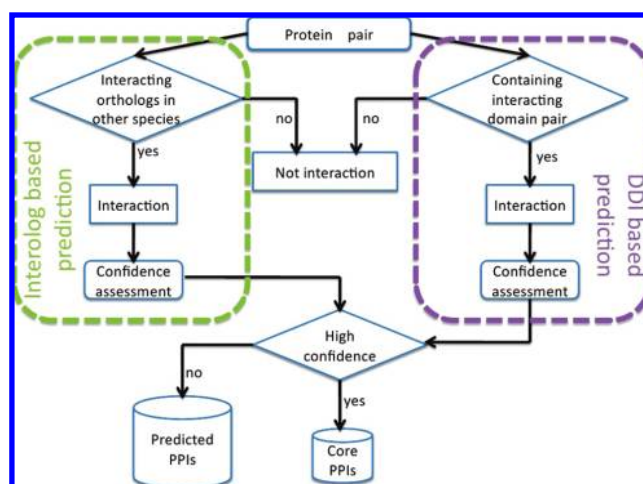


Figure 1. Schematic overview of predicting *F. graminearum* protein–protein interactions based on interologs and domain–domain interactions.

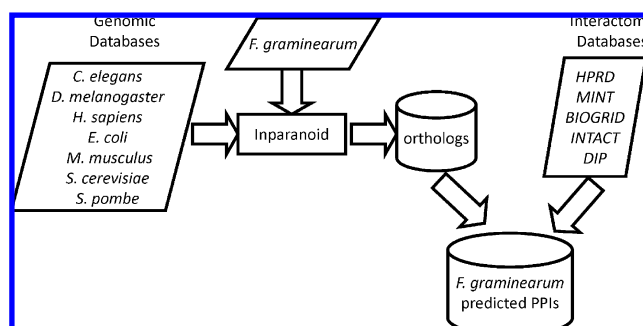


Figure 2. Flowchart for predicting *F. graminearum* protein interactions based on the interolog approach.

and domain–domain interactions. The details of each method will be addressed in the following parts.

Predicting Protein–Protein Interactions for *F. graminearum* Based on Interologs. The approach to predict PPIs based on interaction-ortholog (interolog) is widely used in the literature. In this work, we identified *F. graminearum* protein orthologous from seven well-studied species, including *S. cerevisiae*, *D. melanogaster*, *C. elegans*, *H. sapiens*, *Mus musculus*, *Schizosaccharomyces pombe* and *Escherichia coli* by utilizing INPARANOID.²⁰ The sequence data for *F. graminearum* were downloaded from Broad Institute (http://www.broad.mit.edu/annotation/genome/fusarium_group/MultiDownloads.htm), while the sequence data for other species were obtained from Ensembl²¹ except *S. pombe* and *E. coli*, which were downloaded from NCBI (<ftp://ftp.ncbi.nih.gov/>). In particular, only individual pairs of orthologues from each family were chosen. Although this may omit potential interactions, many more false positives especially those involving divergent inparalogs are discarded. Interactome data sets for the seven organisms were obtained from HPRD (8-22-2007 release),²² MINT (4-8-2008 release),²³ BIOGRID (version 2.0.44),²⁴ INTACT (2008 release)²⁵ and DIP (10-14-2008 release).²⁶ Figure 2 shows the flowchart of predicting protein interactions by the interolog approach. Since interactome data and ortholog data involve different gene identifiers, all the identifiers were mapped to UniProt identifiers, whereas those proteins that do not have any match in UniProt were regarded as not verified and were discarded. *F. graminearum* proteins

were mapped to the collected interactome data through orthologs in other species, where a pair of interacting proteins in a reference species have orthologs in *F. graminearum*. In such a way, the PPIs or a PPI network for *F. graminearum* can be obtained based on the interolog approach.

Predicting Protein–Protein Interactions for *F. graminearum* Based on Domain–Domain Interactions. Despite the success of various computation methods on numerically predicting domain–domain interactions, the rates of false positives and false negatives are very high.²⁷ To alleviate this problem, in this work, only the DDIs experimentally determined based on protein structures were used to predict PPIs, where the DDIs were extracted from iPfam²⁸ and 3did.²⁹ With DDIs available, we can determine whether a pair of proteins interact through following probabilistic model:

$$P_r(P_{ij} = 1) = 1 - \prod_{d_{mn} \in P_{ij}} (1 - P_r(d_{mn} = 1)) \quad (1)$$

Where $P_r(\cdot)$ means the probability of (\cdot) , $P_{ij} = 1$ means that protein i interacts with protein j , $d_{mn} \in P_{ij}$ means that domain m lies in protein i while domain n lies in protein j , and $d_{mn} = 1$ means that domain m interacts with domain n . The idea behind eq 1 is that protein i interacts with protein j as long as there exists a pair of interacting domains in protein pair (i, j) , that is, $P_r(d_{mn} = 1) = 1$. Since the DDIs collected in this work are just annotated with either interaction or noninteraction (i.e., $P_r(d_{mn} = 1) \in \{0, 1\}$), the probability that the corresponding protein pair (i, j) interact is either 0 or 1, that is, $P_r(P_{ij} = 1) \in \{0, 1\}$. Therefore, no threshold is needed to determine whether protein pair (i, j) interact.

To get the domain–domain interactions for *F. graminearum*, we first downloaded gene sequence data from Broad Institute (http://www.broad.mit.edu/annotation/genome/fusarium_group/MultiDownloads.html), where genes have been mapped to Pfam-A domains from Pfam database.²⁸ With the domains available for each gene, a pair of proteins (or corresponding genes) are assumed to interact with each other according to eq 1 if they contain a pair of interacting domains.

Protein–Protein Interaction Confidence. To make it easy for biologists to choose predicted PPIs with high confidence, each PPI is assigned a score representing the interaction confidence. For PPIs predicted by interologs, it is more convincing if the interaction is predicted by more different interactome data sets and interologs in more species. The confidence score S for each pair of interacting proteins predicted by interologs is defined as follows:

$$S_{ij} = N_{\text{interactome}} N_{\text{species}} \quad (2)$$

where S_{ij} is the confidence score for interacting protein pair (i, j) , $N_{\text{interactome}}$ is the number of interactome data sets that support the predicted PPI, and N_{species} is the number of species in which the predicted PPI occurs. The predicted interactions are divided into low confidence ($S_{ij} = 1$), medium confidence ($1 < S_{ij} < 4$) and high confidence ($S_{ij} \geq 4$). The predicted PPIs with low confidence may contain false positives that are observed by some high-throughput techniques such as yeast two-hybrid. Since false positives are seldom observed in multiple species and different interactome data, the PPIs with medium confidence will contain few false positives, whereas the PPIs supported by multiple species and interactome data

sets are more reliable, and therefore PPIs with score $S_{ij} \geq 4$ are regarded as high confident.

For PPIs predicted based on DDIs, the predicted PPIs are classified into three groups, that is, low confidence, medium confidence and high confidence groups. For a pair of interacting proteins, it is regarded as medium-confidence if each protein in the pair contains only one domain and the domain pair involved in the protein pair is an interacting pair. Especially, the interaction is assumed to be high-confidence if more than 50% of sequence length is covered with the domain sequence in each protein of medium-confident interactions. Otherwise, the interacting protein pair is regarded as low-confident because the proteins in the pair may contain multiple domains or motifs, and PPI may be not mediated by DDI, for example, domain–motif interaction.

Verification of Predicted PPIs. To verify the reliability of the predicted PPIs, we examined the extent of coexpression for the predicted PPIs, where a pair of interacting proteins are generally assumed to be coexpressed. Coexpression of a pair of proteins is determined using Pearson Correlation Coefficients (PCCs) calculated based on gene expression data. The microarray data obtained from *F. graminearum* Affymetrix GeneChip³⁰ were downloaded from Plant Expression Database (PLEXdb, <http://www.plexdb.org/index.php>), which is a unified public resource for gene expression data of plants and plant pathogens. In particular, the expression data collected in vitro were used in this work, including expression profiles measured in carbon and nitrogen starvation conditions, in vitro sexual development, conidia germination stages, and in vitro sexual development of *Fusarium* Cch1 calcium channel deletion mutant. To see whether the predicted PPIs' coexpressions are enriched against random pairs, the random pairs were generated by randomly shuffling the links among proteins in predicted PPI network while reserving the degree for each node. Coefficients for the predicted PPIs were calculated and compared against those for the random pairs.

In addition, *F. graminearum* annotations were downloaded from MIPS FGDB, where a small number of genes have been annotated with different functions. Generally, a pair of interacting proteins are assumed to share similar functions through “Guilt by association” rule.³¹ Therefore, function enrichment analysis can verify interactions to some extent. The function enrichment was investigated for the predicted PPIs against random pairs generated as described above. We compared the proportion of PPIs sharing at least one function term in the predicted network against that for random protein pairs.

A straightforward way to verify the predicted interactions is to conduct biological experiments in lab. We randomly selected a set of protein interaction pairs from our predictions, and performed yeast two-hybrid assay to test the interactions in vivo.³² The full-length ORF regions of the selected genes were amplified from genomic DNA extracted from the *F. graminearum* PH-1 strain, and in-frame inserted into pBD-GAL4 and pAD-GAL4 vectors of yeast two-hybrid system (Stratagene, La Jolla, CA). Individual pairs of pAD and pBD plasmids were transformed into yeast PJ69–4A strain and plated on synthetic complete medium without Leu and Trp for transformants. The transformants were titrated on synthetic complete medium without Leu, Trp, His and Ade for interaction tests. The interactions were further confirmed by LacZ assay with X-gal.

Table 1. The Predicted Interactions Based on Interologs from Seven Species

species	predicted interactions
<i>M. musculus</i>	272
<i>C. elegans</i>	723
<i>E. coli</i>	1138
<i>D. melanogaster</i>	1669
<i>S. pombe</i>	2350
<i>H. sapiens</i>	4056
<i>S. cerevisiae</i>	35697

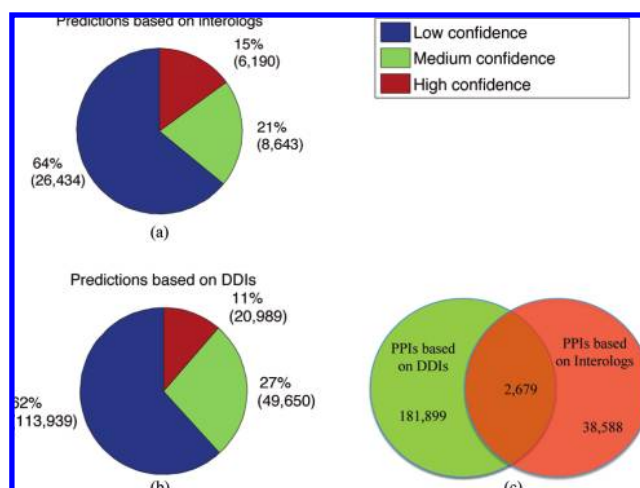
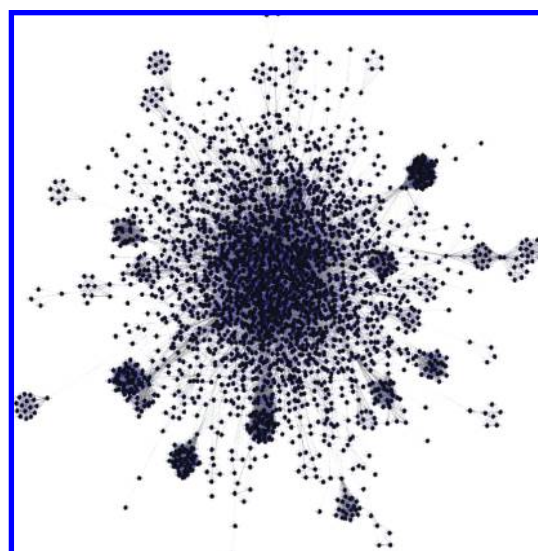
Results and Discussion

***F. graminearum* PPI Map.** Protein–protein interactions were predicted using interolog approach based on the assumption that evolutionarily conserved proteins tend to have conserved interactions. The *F. graminearum* orthologs were respectively identified in *S. cerevisiae*, *D. melanogaster*, *C. elegans*, *H. sapiens*, *M. musculus*, *S. pombe*, and *E. coli* by utilizing INPARANOID. A pair of proteins was regarded as an interacting pair if their corresponding orthologs interact in one of the seven species. Consequently, the predicted PPI data from different species were combined into a PPI network for *F. graminearum*. In this work, 41 267 interactions among 3241 proteins were obtained based on interolog approach. Table 1 lists the interactions predicted from different species.

It can be seen that most of predictions are from yeast (86.50%), which is reasonable from perspective of evolution considering that yeast is close to *F. graminearum* in the phylogeny tree.³³ The predictions from *M. musculus* are much less than those from other species due to the limited number of PPIs available for *M. musculus*. Furthermore, we established interaction confidence for each prediction based on the number of different species from which the interaction was predicted and the number of databases in which the interaction occurs. With the assessment, the predicted PPI data were classified into high-confidence, low-confidence and medium-confidence. Figure 3a shows the predictions with different confidence by interologs. We can see that most of predictions are low-confident, that is, most of predictions are inferred from only one species.

Since interolog approach identifies only conserved interactions across species, the coverage of predicted interactome is generally small. With protein-domain mapping and DDIs from iPfam and 3did, we predicted PPIs with DDIs based on the assumption that PPIs are mediated by DDIs. In this work, 5634 DDIs among 3663 Pfam-A domains were utilized. It has been found that some proteins comprise multiple domains, where the interactions among these proteins may be accomplished through multidomain interactions instead of single interacting domain pair^{18,27} or even by domain–motif interactions like those described in 3did. In this case, the protein interactions predicted based on DDIs possibly contain false positives. On the other hand, we have high confidence in proteins that contain only one domain, where the protein interactions are more possibly accomplished through the corresponding DDIs. As a consequence, the PPIs predicted based on DDIs were classified into low-confidence, medium-confidence and high-confidence sets as described in Methods. Figure 3b shows the predictions with different confidence based on DDIs. From the results, we can see that most predictions are low-confidence.

By integrating predictions by interologs and DDIs, we finally obtained 223 166 interactions among 7406 proteins in total, and

**Figure 3.** The statistics of predicted protein interactions based on DDIs and interologs.**Figure 4.** The protein interaction network for the core data set visualized in Cytoscape.

each protein has 30 partners in average. Figure 3c shows the Venn diagram between PPIs predicted based on interologs and those based on DDIs. It can be seen that the overlap between PPIs predicted with distinct approaches is very small, and therefore, the two approaches can complement each other. With the definition of interaction confidence for predictions, we further constructed a core PPI set that consists of only high-confidence interactions, where the interaction is highly confident if it is predicted as high-confidence by either interologs or DDIs. Furthermore, the interactions that can be predicted by both DDIs and interologs are also regarded as high-confidence even though they are not high-confidence in either DDI based predictions or interolog based predictions. As a consequence, the core PPI data set consists of 27 102 interactions among 3745 proteins and each protein has 7.24 partners in average. Figure 4 shows the network of core PPI data visualized in Cytoscape,³⁴ where only the giant connected component is shown for clearness. It can be seen from the core PPI network that most proteins have few interaction partners while a small number of proteins have a large number of interactions.

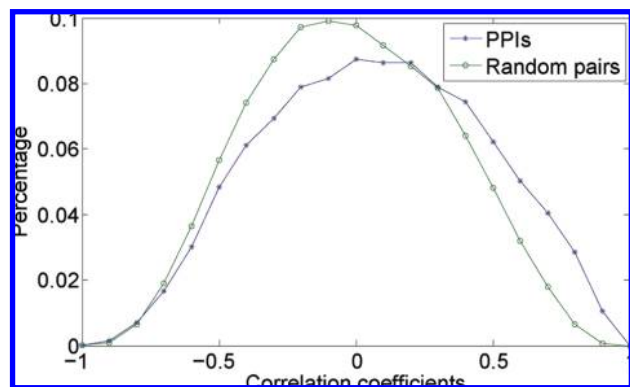


Figure 5. Distributions of correlation coefficients for PPIs from core data set and random protein pairs, respectively.

Verification of Predicted PPIs. To validate the predicted interactome, we investigated whether an interacting protein pair coexpress based on the assumption that interacting proteins tend to have similar gene expression profiles. Note that the lack of coexpression does not necessarily mean that the predicted PPIs do not exist in practice. The coexpression was assessed by Pearson Correlation Coefficients (PCCs), where a high PCC means that the corresponding gene pair is coexpressed and strengthens the confidence of the predicted PPI. We calculated PCCs for predicted interactome against those for random network, which was generated by shuffling links in the predicted PPI network while reserving the degree distribution. Figure 5 shows the distribution of PCCs for PPIs from core data set against those for random protein pairs. It can be learned from the results that most predicted PPIs have higher PCCs than random protein pairs, and this is much clearer for higher correlation coefficients close to 1, which demonstrates that our predicted PPIs are more reliable than random protein pairs.

Furthermore, to verify our predicted PPIs, we investigated the functions of proteins involved in interactions with the assumption that interacting proteins generally have similar functions. Right now, there are 4321 proteins that have been annotated in MIPS FGDB, where 2975 proteins involved in the predicted PPIs have been annotated. To see whether the predicted PPI pair tends to have similar functions, we defined a function similarity score for a pair of proteins:

$$s(i, j) = \frac{|f(i) \cap f(j)|}{|f(i) \cup f(j)|} \quad (3)$$

where $s(i, j)$ is the function similarity score for interacting protein pair (i, j) , $f(i)$ represents the set of functional terms from FGDB for protein i , \cap is the intersection of two sets, \cup is the union of two sets, and $|\cdot|$ means the number of elements in the set, that is, cardinality of the set. The higher $s(i, j)$ is, the higher confidence the interaction pair (i, j) has. The predicted PPIs were compared against random protein pairs which were generated by shuffling links in the predicted PPI network while reserving the degree distribution. Figure 6 shows the distribution of functional similarities for interacting proteins in core data set against those for random protein pairs, where percentage means the percentage of PPIs that have similar functions with similarity score $s(i, j)$. It can be seen from Figure 6 that the predicted PPIs more likely have similar functions than random ones, which verifies the reliability of our predicted PPIs.

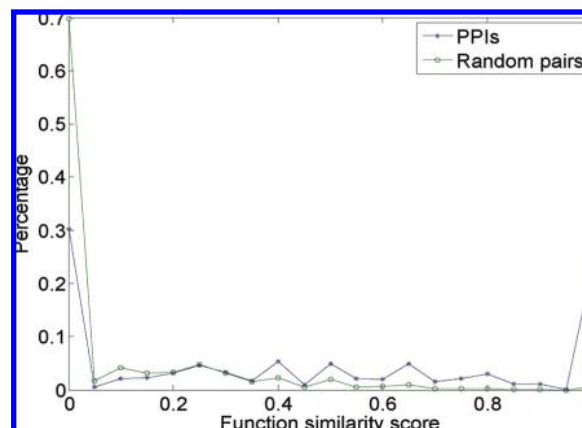


Figure 6. Distributions of function similarity for PPIs from core data set and random protein pairs, respectively.

To have a taste test of experimental verification of the predicted PPIs, we randomly selected two hubs, that is, FGSG_10313 and FGSG_06385. We further selected 12 pairs of proteins related to these two hubs, where 3 pairs occur in core PPI data set. The 12 pairs of interactions are FGSG_10313 and FGSG_00306, FGSG_10313 and FGSG_09229, FGSG_10313 and FGSG_07404, FGSG_10313 and FGSG_06103, FGSG_10313 and FGSG_10822, FGSG_10313 and FGSG_10963, FGSG_06385 and FGSG_00306, FGSG_06385 and FGSG_05535, FGSG_06385 and FGSG_09778, FGSG_06385 and FGSG_10822, FGSG_06385 and FGSG_10963, FGSG_06103 and FGSG_07404. Among the 12 predicted pairs, three pairs, that is, FGSG_10313 and FGSG_07404, FGSG_06385 and FGSG_10822, and FGSG_06103 and FGSG_07404, belong to our core PPI data set. We verified the randomly selected predictions by yeast two-hybrid assay, which is a widely used method for testing protein–protein interactions in vivo. Among the randomly selected 12 predictions, one pair of interaction (FGSG_06103 and FGSG_7404) shows clearly interaction (see Figure 7), which is one of the three pairs in the core PPI data set. FGSG_06103 is a probable phosphoprotein phosphatase 3- α catalytic chain, while FGSG_07404 is a probable calcineurin B, which can be a regulatory unit for phosphatase. Our theoretical and experimental results revealed that this particular pair of phosphatase and calcineurin B interact with each other. Note that the verified interaction pair is among the three randomly selected pairs from our core data set. Although one of 12 randomly selected prediction pairs is verified, one out of 3 predictions from the core PPI data set is proved to be true, which demonstrates that our PPI prediction does provide useful leads for experimental tests. Besides the verified interaction pair, one pair of interactions also predicted in the core data sets is the interaction between FGSG_07404 (the probable calcineurin B) and FGSG_10313, which is MGVI, a mitogen-activated protein (MAP) kinase required for plant infection of *F. graminearum*.³⁵ It has been reported in a human fungal pathogen *Cryptococcus neoformans* that the calcineurin pathway and cell wall integrity MAP kinase Mpk1 pathway cooperate in regulating responses to high temperature stress,³⁶ while the linking molecules between these two pathways has not been identified. Recently in mammal, a calcineurin B has been found in the same complex with a MAP kinase ERK, probably through indirect interaction.³⁷ Therefore, it is possible that the calcineurin B (FGSG_07404) interacts with the MAP kinase MGVI (FGSG_10313) in *F. graminearum*. Another pair of interaction from the core PPI data set is the interaction

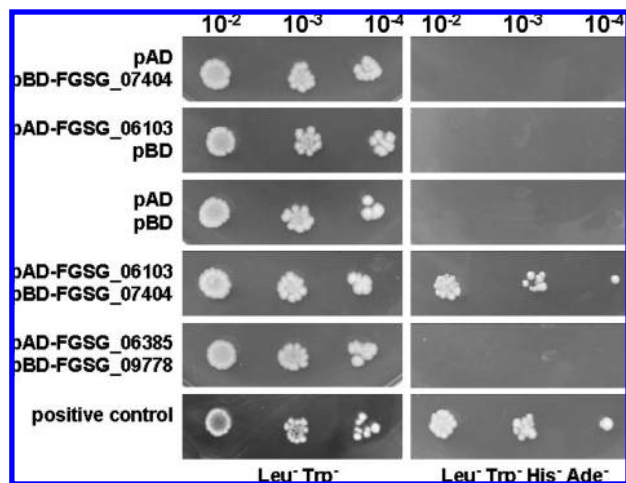


Figure 7. Yeast two-hybrid testing the predicted interactions. (Left panels) The growth of yeast transformants harboring the indicated plasmids on synthetic complete medium without Leu and Trp, which indicates the transformants harbouring pAD and pBD plasmids. (Right panels) The same transformants with same dilution as the left panel were placed on synthetic complete medium without Leu, Trp, Ade and His. The colonies can only grow if there is an interaction between tested protein pairs. The last row shows the positive control.

between FGSG_06385, which is MAP1/GPMK1, a MAP kinase required for pathogenicity^{38,39} and FGSG_10822, which is annotated as a probable Gim complex component GIM3 (GIM represents genes involved in microtubule biogenesis). FGSG_10822 is annotated as a probable GIM3 which belongs to prefoldin family function as chaperones. On the basis of a recent report identifying proteins associated with all known 63 chaperones with TAP-tag in *S. cerevisiae*,⁴⁰ the yeast Gim3 is among the groups of promiscuous chaperones functioning mainly in cytosol, a complex with more than 200 proteins which including some protein kinases. Therefore, it is not surprising that FGSG_10822 might interact with MAP1 (FGSG_06385) in *F. graminearum*. More information from literature about proteins involved in our predictions can be found in Supplementary Table 2. Because we used a stringent condition in yeast two-hybrid, it is possible that we may miss weak or transient interactions in our positive lists. In fact, we did find that all the yeast transformants harbouring testing protein pairs can grow on synthetic medium without Leu, Trp, Ade, which is less stringent condition for interaction tests (see Supplemental Table 1).

Through coexpression analysis and function investigation, we can see that the predicted interacting proteins tend to express together and more likely have similar functions compared against random protein pairs. Although there are only limited gene expression and function annotation data available for verification of our predicted PPIs, the comparison of predicted PPIs against random ones clearly shows that the predicted PPIs are reliable and ready for use. Furthermore, yeast two-hybrid experiments on 12 randomly selected prediction pairs verified one predicted interaction, which is among 3 pairs of interactions from the core PPI data set and proves the reliability of our predictions.

Usage of the FPPI Database. We built an interactive Web server for the FPPI database, which can be accessed through the following link: <http://csb.shu.edu.cn/fppi>. Figure 8 shows the interface of FPPI database. FPPI database contains all

predicted protein interactions for *F. graminearum*, and the core PPI data set with high confidence. In addition, FPPI database also contains functional information for this fungus, including gene information, such as gene sequence, upstream 1k sequence, downstream 1k sequence, and protein sequence information. Furthermore, the database includes the best hits of *F. graminearum* proteins found by BLAST in other databases, including NCBI nonredundant sequence database (RefSeq),⁴¹ KOGS database,⁴² UniProt database⁴³ and MEROPS database,⁴⁴ where the corresponding *P*-values and descriptions for the best hits are available, and the homology information can provide insight into functions of *F. graminearum* genes. Especially, we listed the hierarchical function definitions for the best hits of *F. graminearum* proteins from general functions to more specific functions. In addition, FPPI database contains homology information of *F. graminearum* proteins in other fungi, including *F. graminearum*, *Fusarium verticillioides*, *Fusarium oxysporum*, *Ashbya gossypii*, *Ustilago maydis*, *Phytophthora sojae*, *Neurospora crassa*, *Aspergillus nidulans*, and *S. cerevisiae*. In particular, the number of hits of each *F. graminearum* protein in other fungi was listed with respect to different *E*-value thresholds by BLAST so that it will be easy for biologists to see the evolutionary information among different fungi, which provides valuable hints on pathogenicity of *F. graminearum*.

Conclusions

In this paper, an interactome map was constructed for *F. graminearum* based on evolutionarily conserved interactions across species and domain–domain interactions experimentally determined based on protein structure. Each predicted interaction is assigned a confidence score based on the domain interaction information and the number of organisms that support it. To evaluate the reliability of predicted PPIs, we investigated whether the predicted interacting proteins tend to be coexpressed. With in vitro gene expression data for *F. graminearum* under different conditions, the predicted interacting proteins were found and indeed tend to be coexpressed compared with random ones. Furthermore, we investigated the functions of interacting proteins to see whether interacting proteins tend to have similar functions. The annotations from MIPS FGDB for *F. graminearum* demonstrate that our predicted interacting protein pairs more likely have similar functions compared with random pairs. Although the predicted interactome is far from complete and there are possible false positives in it, the predicted interactome can provide insights into both functions of *F. graminearum* proteins and the mechanism underlying pathogenicity genes.

In addition, we performed yeast two-hybrid experiments to verify a small set of interactions randomly selected from our predictions. Twelve pairs of interactions were tested, where three experiments are designed for each interaction. All 12 pairs of interactions passed at least one experiment and therefore show at least weak interactions, whereas only one pair of interaction (from the core PPI data set) passed all three experiments and is regarded as real interaction, which verified the effectiveness of the proposed method. Although only a small set of randomly selected predictions were tested in lab, the results prove the reliability of the predicted interactome to some extent. In addition, information from literature show that some other predictions are possibly real interactions. FPPI also contains other functional information for *F. graminearum* genes, including homologues in other species deposited in

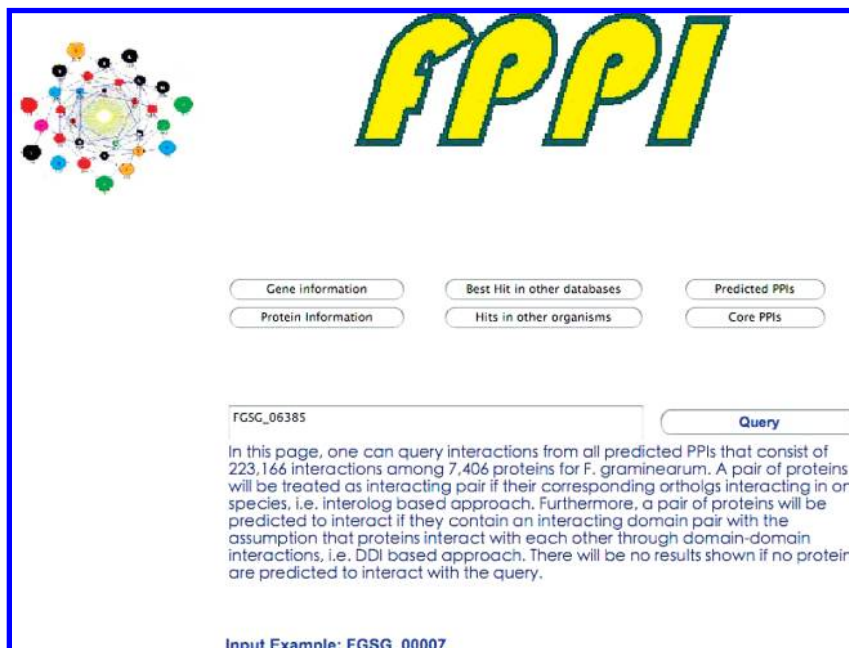


Figure 8. The interface of FPPI database.

different databases and the inferred functional characteristics, and so on. In summary, FPPI is a rich source of information for system-level understanding of gene function and biological processes in *F. graminearum*.

Acknowledgment. This paper was partly supported by Innovation Funding of Shanghai University, Knowledge Innovation Program of the Chinese Academy of Sciences (Grant No. KSCX2-YW-N-005), Innovation Program of Shanghai Municipal Education Commission (09ZZ93), Open Funding of National Key Laboratory of Plant Molecular Genetics, and SRF for ROCS, SEM.

Supporting Information Available: The information about Y2H experiments for 12 protein pairs is available in Supplementary Table 1. Related reference information about 9 predictions can be found in Supplementary Table 2. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References

- Goswami, R.; Kistler, H. Heading for disaster: *Fusarium graminearum* on cereal crops. *Mol. Plant Pathol.* **2004**, *5*, 515–525.
- Brewing Microbiology*; Priest, F., Campbell, I., Eds.; Springer: New York, 2002; Vol. 3.
- Winnenburg, R.; Baldwin, T. K.; Urban, M.; Rawlings, C.; Kohler, J.; Hammond-Kosack, K. E. PHI-base: a new database for pathogen host interactions. *Nucleic Acids Res.* **2006**, *34*, D459–464.
- Gohre, V.; Robatzek, S. Breaking the barriers: microbial effector molecules subvert plant immunity. *Annu. Rev. Phytopathol.* **2008**, *46*, 189–215.
- Cuomo, C. A.; et al. The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. *Science* **2007**, *317*, 1400–1402.
- Guldener, U.; Mannhaupt, G.; Munsterkotter, M.; Haase, D.; Oesterheld, M.; Stumpflen, V.; Mewes, H.-W.; Adam, G. FGDB: a comprehensive fungal genome resource on the plant pathogen *Fusarium graminearum*. *Nucleic Acids Res.* **2006**, *34*, D456–458.
- Uetz, P.; et al. A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* **2000**, *403*, 623–627.
- Li, S.; et al. A map of the interactome network of the metazoan *C. elegans*. *Science* **2004**, *303*, 540–543.
- Giot, L.; et al. A protein interaction map of *Drosophila melanogaster*. *Science* **2003**, *302*, 1727–1736.
- Uetz, P.; Pankratz, M. J. Protein interaction maps on the fly. *Nat. Biotechnol.* **2004**, *22*, 43–44.
- Rual, J.-F.; et al. Towards a proteome-scale map of the human protein-protein interaction network. *Nature* **2005**, *437*, 1173–1178.
- Chen, L.; Wang, R.-S.; Zhang, X.-S. *Biomolecular Networks: Methods and Applications in Systems Biology*; John Wiley & Sons: Hoboken, NJ, 2009.
- Matthews, L.; Vaglio, P.; Reboul, J.; Ge, H.; Davis, B.; Garrels, J.; Vincent, S.; Vidal, M. Identification of potential interaction networks using sequence-based searches for conserved protein-protein interactions or “interologs”. *Genome Res.* **2001**, *11*, 2120–2126.
- Ideker, T.; Ozier, O.; Schwikowski, B.; Siegel, A. Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics* **2002**, *18*, S233–240.
- Wu, X.; Zhu, L.; Guo, J.; Zhang, D.; Lin, K. Prediction of yeast protein-protein interaction network: insights from the Gene Ontology and annotations. *Nucleic Acids Res.* **2006**, *34*, 2137–2150.
- Deng, M.; Mehta, S.; Sun, F.; Chen, T. Inferring domain-domain interactions from protein-protein interactions. *Genome Res.* **2002**, *12*, 1540–1548.
- Chen, L.; Wu, L.-Y.; Wang, Y.; Zhang, X.-S. Inferring protein interactions from experimental data by association probabilistic method. *Proteins* **2006**, *62*, 833–837.
- Wang, R.-S.; Wang, Y.; Wu, L.-Y.; Zhang, X.-S.; Chen, L. Analysis on multi-domain cooperation for predicting protein-protein interactions. *BMC Bioinf.* **2007**, *8*, 391.
- Jothi, R.; Kann, M.; Przytycka, T. Predicting protein-protein interaction by searching evolutionary tree automorphism space. *Bioinformatics* **2005**, *21*, i241–250.
- Berglund, A.-C.; Sjolund, E.; Ostlund, G.; Sonnhammer, E. L. L. InParanoid 6: eukaryotic ortholog clusters with inparalogs. *Nucleic Acids Res.* **2008**, *36*, D263–266.
- Hubbard, T. J. P.; et al. Ensembl 2009. *Nucleic Acids Res.* **2009**, *37* (Database issue), D690–697.
- Mishra, G. R.; et al. Human protein reference database-2006 update. *Nucleic Acids Res.* **2006**, *34*, D411–414.
- Chatr-aryamontri, A.; Ceol, A.; Palazzi, L. M.; Nardelli, G.; Schneider, M. V.; Castagnoli, L.; Cesareni, G. MINT: the Molecular INTeraction database. *Nucleic Acids Res.* **2007**, *35*, D572–574.
- Breitkreutz, B.-J.; Stark, C.; Reguly, T.; Boucher, L.; Breitkreutz, A.; Livstone, M.; Oughtred, R.; Lackner, D. H.; Bahler, J.; Wood, V.; Dolinski, K.; Tyers, M. The BioGRID Interaction Database: 2008 update. *Nucleic Acids Res.* **2008**, *36*, D637–640.
- Kerrien, S.; et al. IntAct-open source resource for molecular interaction data. *Nucleic Acids Res.* **2007**, *35*, D561–565.

- (26) Salwinski, L.; Miller, C. S.; Smith, A. J.; Pettit, F. K.; Bowie, J. U.; Eisenberg, D. The Database of Interacting Proteins: 2004 update. *Nucleic Acids Res.* **2004**, *32*, D449–451.
- (27) Zhao, X.-M.; Chen, L. Domain-Domain Interaction Identification with a Feature Selection Approach. *Lecture Notes in Bioinformatics. Proceedings of the Third IAPR International Conference on Pattern Recognition in Bioinformatics*; Springer-Verlag: Berlin, Heidelberg, 2008; pp 178–186.
- (28) Finn, R. D.; Marshall, M.; Bateman, A. iPfam: visualization of protein-protein interactions in PDB at domain and amino acid resolutions. *Bioinformatics* **2005**, *21*, 410–412.
- (29) Stein, A.; Russell, R. B.; Aloy, P. 3did: interacting protein domains of known three-dimensional structure. *Nucleic Acids Res.* **2005**, *33*, D413–417.
- (30) Guldener, U.; Seong, K.-Y.; Boddu, J.; Cho, S.; Trail, F.; Xu, J.-R.; Adam, G.; Mewes, H.-W.; Muehlbauer, G. J.; Kistler, H. C. Development of a *Fusarium graminearum* Affymetrix GeneChip for profiling fungal gene expression in vitro and in planta. *Fungal Genet. Biol.* **2006**, *43*, 316–325.
- (31) Zhao, X.-M.; Wang, Y.; Chen, L.; Aihara, K. Gene function prediction using labeled and unlabeled data. *BMC Bioinf.* **2008**, *9*, 57.
- (32) Tang, W.; Ezcurra, I.; Muschietti, J.; McCormick, S. A cysteine-rich extracellular protein, LAT52, interacts with the extracellular domain of the pollen receptor kinase LePRK2. *Plant Cell* **2002**, *14*, 2277–2287.
- (33) Fitzpatrick, D.; Logue, M.; Stajich, J.; Butler, G. A fungal phylogeny based on 42 complete genomes derived from supertree and combined gene analysis. *BMC Evol. Biol.* **2006**, *6*, 99.
- (34) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13*, 2498–2504.
- (35) Hou, Z.; Xue, C.; Peng, Y.; Katan, T.; Kistler, H. C.; Xu, J.-R. A mitogen-activated protein kinase gene (MGV1) in *Fusarium graminearum* is required for female fertility, heterokaryon formation, and plant infection. *Mol. Plant–Microbe Interact.* **2002**, *15*, 1119–1127, PMID: 12423017.
- (36) Kraus, P.; Fox, D.; Cox, G.; Heitman, J. The *Cryptococcus neoformans* MAP kinase Mpk1 regulates cell integrity in response to antifungal drugs and loss of calcineurin function. *Mol. Microbiol.* **2003**, *48*, 1377–387.
- (37) Dougherty, M. K.; Ritt, D. A.; Zhou, M.; Specht, S. I.; Monson, D. M.; Veenstra, T. D.; Morrison, D. K. KSR2 is a calcineurin substrate that promotes ERK cascade activation in response to calcium signals. *Mol. Cell* **2009**, *34*, 652–662.
- (38) Cuzick, A.; Urban, M.; Hammond-Kosack, K. *Fusarium graminearum* gene deletion mutants map1 and tri5 reveal similarities and differences in the pathogenicity requirements to cause disease on Arabidopsis and wheat floral tissue. *New Phytol.* **2008**, *177*, 990–1000.
- (39) Jenczmionka, N. J.; Maier, F. J.; Löscher, A. P.; Schäfer, W. Mating, conidiation and pathogenicity of *Fusarium graminearum*, the main causal agent of the head-blight disease of wheat, are regulated by the MAP kinase gpmk1. *Curr. Genet.* **2003**, *43*, 87–95.
- (40) Gong, Y.; Kakiyama, Y.; Krogan, N.; Greenblatt, J.; Emili, A.; Zhang, Z.; Houry, W. A. An atlas of chaperone-protein interactions in *Saccharomyces cerevisiae*: implications to protein folding pathways in the cell. *Mol. Syst. Biol.* **2009**, *5*, 275.
- (41) Pruitt, K. D.; Tatusova, T.; Maglott, D. R. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **2005**, *33*, D501–504.
- (42) Tatusov, R.; et al. The COG database: an updated version includes eukaryotes. *BMC Bioinf.* **2003**, *4*, 41.
- (43) Consortium, T. U. The Universal Protein Resource (UniProt). *Nucleic Acids Res.* **2008**, *36*, D190–195.
- (44) Rawlings, N. D.; Morton, F. R.; Kok, C. Y.; Kong, J.; Barrett, A. J. MEROPS: the peptidase database. *Nucleic Acids Res.* **2008**, *36*, D320–325.

PR900415B