Amino Acids (2012) 42:2363–2371
DOI 10.1007/s00726-011-0978-z

ORIGINAL ARTICLE

Prediction of protein–protein interactions between *Ralstonia solanacearum* and *Arabidopsis thaliana*

Zhi-Gang Li · Fei He · Ziding Zhang · You-Liang Peng

Received: 2 January 2011 / Accepted: 6 July 2011 / Published online: 24 July 2011

Abstract *Ralstonia solanacearum* is a devastating bacterial pathogen that has an unusually wide host range. *R. solanacearum*, together with *Arabidopsis thaliana*, has become a model system for studying the molecular basis of plant–pathogen interactions. Protein–protein interactions (PPIs) play a critical role in the infection process, and some PPIs can initiate a plant defense response. However, experimental investigations have rarely addressed such PPIs. Using two computational methods, the interolog and the domain-based methods, we predicted 3,074 potential PPIs between 119 *R. solanacearum* and 1,442 *A. thaliana* proteins. Interestingly, we found that the potential pathogen-targeted proteins are more important in the *A. thaliana* PPI network. To facilitate further studies, all predicted PPI data were compiled into a database server called PPIRA (http://protein.cau.edu.cn/ppira/). We hope that our work will provide new insights for future research addressing the pathogenesis of *R. solanacearum.*

Z.-G. Li and F. He contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s00726-011-0978-z) contains supplementary material, which is available to authorized users.

Z.-G. Li · F. He · Z. Zhang (✉) · Y.-L. Peng
State Key Laboratory of Agrobiotechnology, China Agricultural University, Beijing 100193, China
e-mail: zidingzhang@cau.edu.cn

Z.-G. Li · Y.-L. Peng
Department of Plant Pathology, China Agricultural University, Beijing 100193, China

F. He · Z. Zhang
Bioinformatics Center, College of Biological Sciences, China Agricultural University, Beijing 100193, China

Keywords Bioinformatics · Pathogenicity · Plant–pathogen interactions · Prediction · Protein–protein interaction

Introduction

Ralstonia solanacearum (previously named *Pseudomonas solanacearum*) is a β-proteobacteria that causes severe worldwide agricultural losses. This gram-negative bacterium can infect plants through the roots. First, it enters the xylem vessel where it grows and reproduces rapidly, leading to wilting disease that causes the host to die. Then, it goes back into the soil and returns to a saprophytic organism. Susceptible crops generally cannot be farmed on infected land for a long period of time (Hayward 1991). *R. solanacearum* has an extraordinarily wide range of hosts, including solanaceae family plants, leguminous plants and some mono-cotyledonous plants. In total, more than 200 species, which cover at least 20 botanical families, are potential hosts. Molecular analysis has revealed that *R. solanacearum* can be divided into five races based on whether the hosts originated from Asia, America or Africa. Unfortunately, effective control is not available for this devastating pathogen (Stéphane and Christian 2002).

Many genes of the *R. solanacearum* bacterium that are responsible for its pathogenesis have been identified after decades of careful work. The genome sequence of *R. solanacearum* was released in 2002 (Salanoubat et al. 2002), allowing us to investigate the mechanism of its pathogenicity at the whole genome scale. This bacterium and one of its hosts (*Arabidopsis thaliana*) have become a model system for studying plant–bacteria interactions because of the genetic and molecular tractability of both the pathogen and the host (Salanoubat et al. 2002). However, the

infection process used by *R. solanacearum* has not been clarified.

Approximately 3,000 proteins in *A. thaliana* are directly related to plant defense (Bishop et al. 2000). Many of these proteins interact directly with the pathogen proteins, and some of them can initiate plant defense responses to the infection. Understanding the protein–protein interaction (PPI) network (i.e., interactome) between plant proteins and pathogen proteins is a critical step for studying the molecular basis of pathogenesis (Pinzon et al. 2011; He et al. 2008; Kim et al. 2008). However, it is still a challenging task to identify the plant proteins targeted by a pathogen protein through existing experimental techniques (Bogdanove 2002). Currently, only a few pairs of such interactions have been identified, which is far from being enough to systematically decipher the molecular mechanism of pathogenicity.

In the past decade, a series of PPI prediction methods have been elegantly developed, and they are playing an increasingly important role in complementing experimental approaches. Diverse data types or properties, such as gene ontology (GO) annotations (Wu et al. 2006), protein sequence similarity (Matthews et al. 2001), protein domain interactions (Ng et al. 2003), and protein structural information (Ogmen et al. 2005), have been frequently utilized to construct PPI prediction methods. Two widely implemented methods are probably the interolog and the domain-based methods (Shoemaker and Panchenko 2007; Florez et al. 2010).

In this work, the PPIs between *R. solanacearum* and *A. thaliana* were jointly predicted using the interolog and the domain-based methods. The interolog method relies on protein sequence similarity to conduct the PPI prediction. Briefly, an *R. solanacearum* protein and an *A. thaliana* protein can be predicted to interact with each other if an experimentally verified interaction exists between their respective homologous proteins in another organism. The domain-based method uses domain interaction information, which is derived from known protein 3D structures, to infer the potential PPIs. If an *R. solanacearum* protein and an *A. thaliana* protein contain an interacting domain pair, we can expect the two proteins to also interact with each other. We present the construction of a prediction pipeline and the detailed results of our PPI prediction. In addition, we discuss how these predicted PPIs can help us to better the understand plant–bacteria interactions.

Materials and methods

Datasets

A total of 5,113 *R. solanacearum* protein sequences were downloaded from the National Center for Biotechnology Information (NCBI) database (ftp://ftp.ncbi.nih.gov/genomes/Bacteria/Ralstonia_solanacearum/). Overall, 28,064 *A. thaliana* protein sequences were obtained from the TIGR database (http://www.tigr.org/tdb/e2k1/ath1/).

To implement the interolog method, 56,191 experimentally verified PPIs were obtained from the Database of Interacting Proteins (DIP; http://dip.doe-mbi.ucla.edu/dip/Main.cgi) (Salwinski et al. 2004). To perform the domain-based PPI prediction, we downloaded the iPfam database (version 23) (http://ipfam.sanger.ac.uk/) (Bateman et al. 2000), which contains 4,025 interacting Pfam domain pairs.

For the interactome of *A. thaliana*, 4,660 experimentally determined PPIs were downloaded from three public databases: The Arabidopsis Information Resource (TAIR) (http://www.arabidopsis.org) (Swarbreck et al. 2008), IntAct (http://www.ebi.ac.uk/intact/main.xhtml) (Kerrien et al. 2007) and BioGrid (http://www.thebiogrid.org) (Stark et al. 2006). The predicted *A. thaliana* PPI dataset was also obtained from TAIR. To ensure that the predicted data are generally reliable, those with a low confidence value (i.e., CV < 2) were discarded.

The identification of secreted and membrane proteins

Membrane proteins in *R. solanacearum* were identified by TMHMM (version 2.0) (Sonnhammer et al. 1998; Krogh et al. 2001), which is considered one of the best transmembrane protein predictors (Moller et al. 2001). The proteins were inferred to be transmembrane if the number of predicted transmembrane helices was not <1, and the expected number of amino acids in at least one transmembrane helix was not <18. SignalP (version 3.0) (Bendtsen et al. 2004) was employed to identify secretory proteins with the strictest criterion.

The prediction of PPIs between *R. solanacearum* and *A. thaliana*

The potential PPIs between *R. solanacearum* and *A. thaliana* were predicted using the interolog method (Fig. 1). Briefly, each protein in *R. solanacearum* or *A. thaliana* was first BLASTed against all the proteins in the DIP database to identify homologs with *E*-value, sequence identity and aligned sequence length coverage cut-offs of 0.001, 30 and 80%, respectively. For each protein pair between *R. solanacearum* and *A. thaliana,* an interaction can be predicted if their corresponding homologs in DIP have at least one interaction.

The potential PPIs between *R. solanacearum* and *A. thaliana* were also predicted using the domain-based method (Fig. 1). The Pfam domain information for *A. thaliana* proteins were obtained from TAIR, whereas

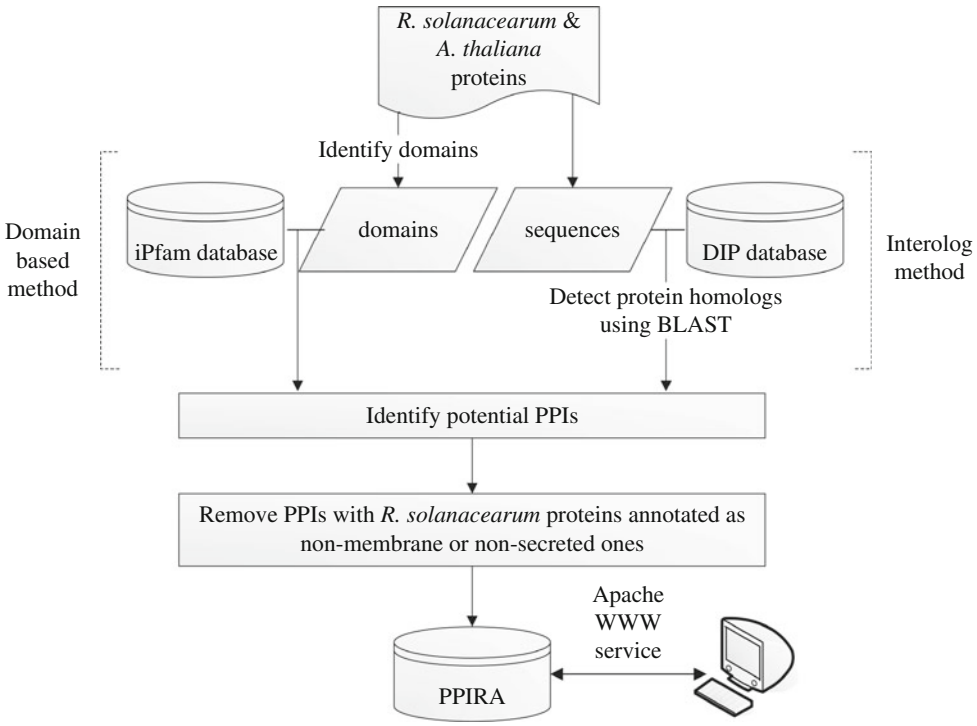

Fig. 1 The prediction pipeline of the potential PPIs between *R. solanacearum* and *A. thaliana*

R. solanacearum proteins were directly submitted to the Pfam server (Bateman et al. 2000) to identify the domains with *E*-value and aligned sequence length coverage cut-offs of 0.001 and 90%, respectively. If a protein pair between *R. solanacearum* and *A. thaliana* contains an interacting Pfam domain pair, the protein pair is expected to interact with each other.

The predicted PPIs based on the interolog and the domain-based methods were merged into a PPI network between *R. solanacearum* and *A. thaliana*. Furthermore, those PPIs with *R. solanacearum* proteins predicted to be non-membrane or non-secreted ones were removed from the established PPI network.

Analysis of GO enrichment

GO annotations of 96% of the *A. thaliana* gene products were obtained from the GO website (http://www.geneontology.org/). GO annotations of the *R. solanacearum* genome were obtained from B2G-FAR (http://bioinfo.cipf.es/b2gfar/home) (Gotz et al. 2008). About 80% of the gene products of *R. solanacearum* were annotated. According to the GO hierarchy, we looked up parent terms with an "is_a" relationship of each GO term to find GO terms at different hierarchies for a gene product. For each group of gene products, we calculated the proportion of each GO term at the sixth level of the GO hierarchy and did a Fisher exact test to determine the *p* value followed by a false discovery rate (FDR) correction (Storey 2002).

Clustering of the *A. thaliana* PPI network

The CFinder program (http://www.cfinder.org, v2.0.1) (Adamcsek et al. 2006), which is based on the clique percolation algorithm (Derenyi et al. 2005), was employed to cluster the *A. thaliana* PPI network. $k = 3$ was used to cluster the whole *A. thaliana* PPI network, whereas $k = 4$ was used to re-cluster the largest cluster generated at $k = 3$. In each of the identified cluster, GO enrichment was determined using the Fisher exact test followed by the FDR correction. If many enriched GO terms existed in a cluster, only the most significantly enriched GO term was assigned.

Network topology analysis

Two topological parameters (i.e., degree and betweenness) for each protein in the *A. thaliana* PPI network were computed using NetworkAnalyzer (Assenov et al. 2008). In a PPI network, each protein is represented as a node. The degree of a node is simply defined as the number of interactions that a node has. The betweenness is a centrality measure of a node in a network. The betweenness centrality of a node n can be computed as follows:

$$C_{\mathrm{b}}(n) = \sum_{s \neq n \neq t} (\sigma_{st}(n)/\sigma_{st})$$

where s and t are nodes in the network that are different from n, σ_{st} denotes the number of shortest paths from s to t, and $\sigma_{st}(n)$ is the number of shortest paths from s to t that pass through the node n.

Results and discussion

The predicted PPIs between *R. solanacearum* and *A. thaliana*

We predicted a total of 3,074 possible interactions between 119 *R. solanacearum* and 1,442 *A. thaliana* proteins, which were compiled into a PPI network called PPIRA. Of the predicted PPIs, 1,438 and 1,648 interactions were predicted by the interolog and the domain-based methods, respectively, whereas 12 interactions were consistently predicted by both the methods. On average, an *R. solanacearum* protein has around 26 *A. thaliana* interacting partners, while an *A. thaliana* protein interacts with only approximately two pathogen proteins (Fig. 2). The ratio of proteins involved in the predicted PPI network between *R. solanacearum* and *A. thaliana* is generally in line with a previous computational work that addressed the interactome between another plant pathogen, *Xanthomonas oryzae*, and rice (Kim et al. 2008). As reported by Kim et al. (2008), the plant pathogen *Xanthomonas oryzae* expresses a few proteins to attack its host's proteome. It has been established that a pathogen mutates its genes extensively to infect a host, whereas a plant defends the attacks by expanding its gene families (Stahl and Bishop 2000). Therefore, to some extent, the ratio of proteins involved in the predicted PPI network may reflect the plant–pathogen arms race at the molecular level.

Our predicted PPI data can be a valuable resource for the community to further study the pathogenesis of bacteria, which can be exemplified by the *R. solanacearum* protein, Pme (NP_521699). As a bacterial virulence protein, Pme is a pectinesterase, which can catalyze the hydrolysis of the thioester bond in palmitoyl-CoA. It participates in cell wall metabolism by catalyzing pectin degradation. Pme may lead to the maceration and cell death of plant tissue, but the detailed infection process is poorly understood (Spok et al. 1991). In the established PPI network (i.e., PPIRA), Pme was predicted to interact with two *A. thaliana* bifunctional dihydrofolate reductase-thymidylate synthases (AT2G16370 and AT4G34570), which are involved in dTMP biosynthesis. Within the context of the established PPI network, this finding provides new clues to further investigate the Pme-related infection process between *R. solanacearum* and *A. thaliana*.

Biological functions of *R. solanacearum* and *A. thaliana* proteins in the established PPI network

To determine if there is a certain biological function bias between the *R. solanacearum* and *A. thaliana* proteins in the established PPI network, we investigated the functional compositions of the corresponding proteins via the analysis of GO enrichment. The over-represented biological functions are different in the two species. The *R. solanacearum* proteins in the predicted PPIs are mainly enriched in

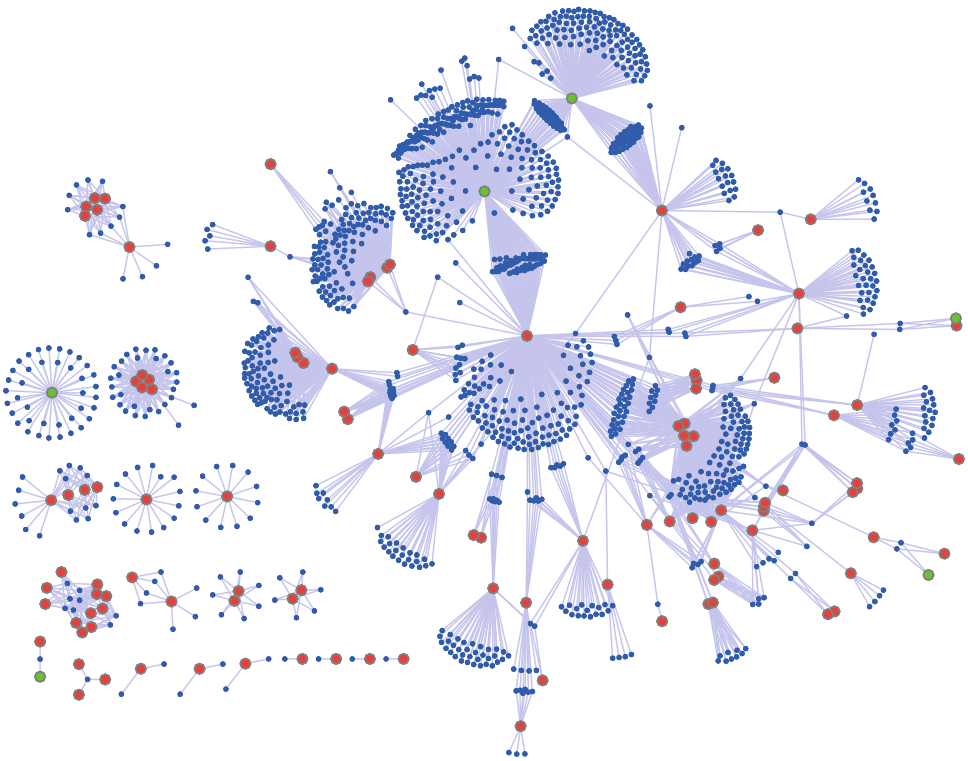

Fig. 2 Graphical representation of PPIs between *R. solanacearum* and *A. thaliana*. Each node represents a protein and each edge denotes an interaction. *Red* and *blue* nodes are *R. solanacearum* and *A. thaliana* proteins, respectively. *Green* nodes are *R. solanacearum* proteins that are encoded by known or candidate genes responsible for pathogenesis (Salanoubat et al. 2002) (color figure online)

transportation proteins (Supplemental File 1). In comparison, the *A. thaliana* proteins predicted to be targeted by the pathogen have diverse functional propensities (Supplemental File 1). Plants evolve various self-protective mechanisms after billions of years of battling against their environment. Whether a plant can survive a war with a pathogen depends on its self-protective systems (Stahl and Bishop 2000). Therefore, proteins with functions that respond to environmental stimuli are enriched in the *A. thaliana* proteins. In addition, the potential pathogen-targeted *A. thaliana* proteins are enriched in signal transduction, regulation and transportation proteins (Supplemental File 1). Such differing functional enrichments between the *A. thaliana* and *R. solanacearum* proteins indicate the different evolutionary traits between plants and pathogens (Stahl and Bishop 2000).

R. solanacearum proteins tend to interact with *A. thaliana* proteins that are more important in the *A. thaliana* PPI network

In a PPI network, a densely connected area is referred to as a cluster, and generally the cluster itself is a functional module. Members of a cluster are usually involved in similar biological processes, and protein complexes can be identified through the clustering of a network (Palla et al. 2005; Jonsson et al. 2006). To investigate the functional modules in which the potential pathogen-targeted *A. thaliana* proteins are involved, we first collected a total of 4,660 experimentally determined *A. thaliana* PPIs from three public databases. Of the 2,292 proteins in the *A. thaliana* PPI network, 265 proteins were predicted to interact with *R. solanacearum.* The CFinder software was then employed to cluster the *A. thaliana* PPI network (Adamcsek et al. 2006). In CFinder, a network can be divided into different *k*-clique clusters. Generally, a larger value of k corresponds to the generation of denser clusters. A total of 83 clusters were generated at $k = 3$, of which 22 contained at least one pathogen-targeted protein. These 22 clusters are referred to as pathogen-targeted clusters in this study.

From a network viewpoint, cellular processes attacked by the pathogen might be revealed from these pathogen-targeted clusters. According to the GO enrichment analysis, we found that biological functions such as regulation of the cell cycle ($p = 3.30 \times 10^{-28}$), channel activity ($p = 5.07 \times 10^{-10}$) and regulation of cellular metabolic processes ($p = 8.89 \times 10^{-16}$) are over-represented in these 22 clusters (Fig. 3a and Supplemental File 2).

There are 52 potential pathogen-targeted proteins in these 22 clusters. Three of them exist in more than one cluster, and these can be considered as the bottleneck of the network and may be involved in the multiple cellular processes. These kinds of proteins usually play vital roles because they likely connect adjacent cellular processes (Yu et al. 2007). For example, CDKA1 was predicted to be targeted by the pathogen proteins. It connects two clusters (#12 and #17 in Fig. 3a) in the *A. thaliana* PPI network. These two clusters are related to biopolymer catabolic processes and the regulation of cell cycles. CDKA1 has previously been identified to function in cell morphogenesis as well as in the cell proliferation, which controls both G1/S and G2/M (mitosis) phase transitions in the cell cycle (Iwakawa et al. 2006). CDKA1 has also been reported to be induced by nematode infection of roots (Niebel et al. 1996). Therefore, the pathogen interacting partners of CDKA1 predicted from our work may be a good starting point to further study the role of this *A. thaliana* protein during the bacterial infection.

Some of the 52 proteins play important roles in the identified network clusters. Here, we take the largest cluster (#9 in Fig. 3a), which contains 148 proteins, as an illustrative example. Cluster #9 can be divided into seven subclusters (Fig. 3b). Most of the proteins that appear in more than one subcluster are calmodulins (CAMs) and calcium-dependent protein kinases (CDPKs), both of which are important in plant innate immune responses (Du et al. 2009; Boudsocq et al. 2010). Interestingly, they are also potential interacting partners of the pathogen proteins. Therefore, potential pathogen-targeted proteins tend to be bottlenecks between the seven subclusters.

The network topological features of the potential pathogen-targeted proteins in the *A. thaliana* interactome can also reflect their biological roles. The *A. thaliana* PPI dataset used in the aforementioned analysis is based only on experimentally determined PPIs and covers just a small part of the full *A. thaliana* interactome (Cui et al. 2008). To better understand the topology of potential pathogen-targeted proteins in a more comprehensive *A. thaliana* interactome, the predicted PPI data obtained from TAIR were also taken into account. Thus, the newly compiled *A. thaliana* PPI network contained 5,240 proteins and 18,196 interactions, and 596 proteins were predicted to interact with the pathogen *R. solanacearum*. In general, these potential pathogen-targeted proteins have a higher degree as well as a larger betweenness than other proteins in the network (Wilcoxon rank-sum test, $p = 0.0069$ and 5.1×10^{-5}). Furthermore, we also observed that the potential pathogen-targeted *A. thaliana* proteins preferred to exist in large clusters at different clustering cut-offs (Wilcoxon rank-sum tests, $p < 0.05$; Fig. 4).

It has been established that human cancer proteins tend to be the bottlenecks of the human PPI network. On average, cancer proteins have more interacting partners than non-cancer proteins, because they generally play roles in more complex cellular processes (Jonsson and Bates 2006). In our analysis, the potential pathogen-targeted *A. thaliana* proteins displayed a similar network topology

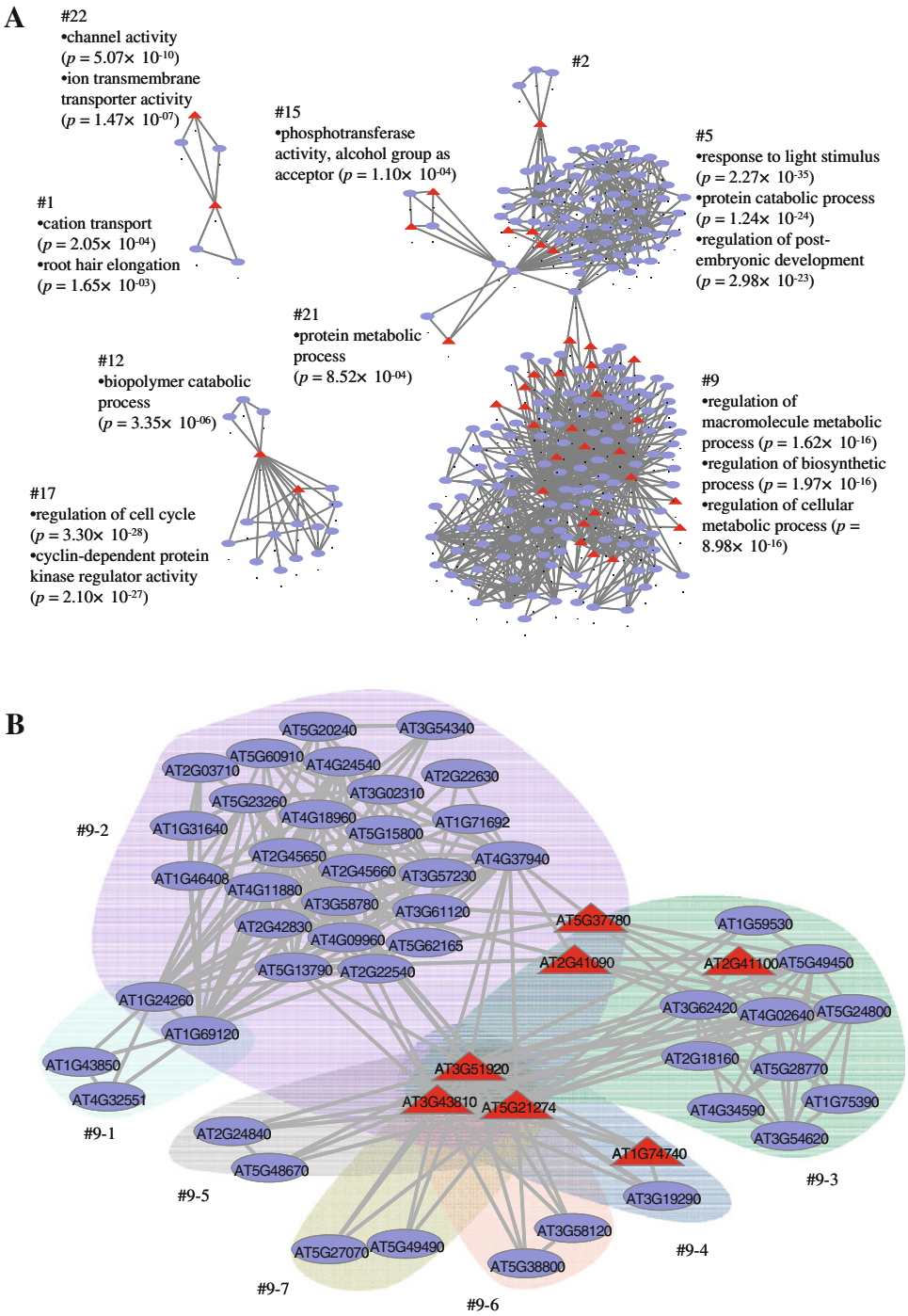

Fig. 3 The identified *A. thaliana* PPI clusters that may be targeted by *R. solanacearum* proteins. Proteins that may interact with *R. solanacearum* are shown as *triangles*. **a** Nine representative clusters ($k = 3$). The GO enrichment of each cluster is also listed. Due to space constraints, the other 13 *A. thaliana* protein clusters that may also be targeted by *R. solanacearum* are not shown. See Supplemental File 2 for detailed information for all 22 protein clusters. **b** Re-clustering ($k = 4$) of the largest cluster generated at $k = 3$ (i.e., cluster #9 in Fig. 3a). The identified subclusters are highlighted in different colors. Most of the proteins that may have an interaction with *R. solanacearum* are shared by different subclusters, indicating they are the bottlenecks in the cluster #9

as cancer proteins in the human interactome, indicating that a successful infection involves the interaction between *R. solanacearum* proteins and *A. thaliana* proteins that play roles in complex biological processes.

The cut-off choice in the PPI prediction methods

Recently, the interolog and the domain-based methods have been used in predicting inter-species PPIs [e.g. the PPI prediction between *Xanthomonas oryzae* and rice (Kim et al. 2008)]. In our work, we employed these two methods using the similar thresholds as Kim et al. (2008) to predict the PPIs between *R. solanacearum* and *A. thaliana*. Comparatively, the interolog and the domain-based methods have been more widely used in predicting intra-species PPIs and the corresponding thresholds have been clearly benchmarked. For instance, Yu et al. (2004) studied the transferability of a PPI between genomes using the

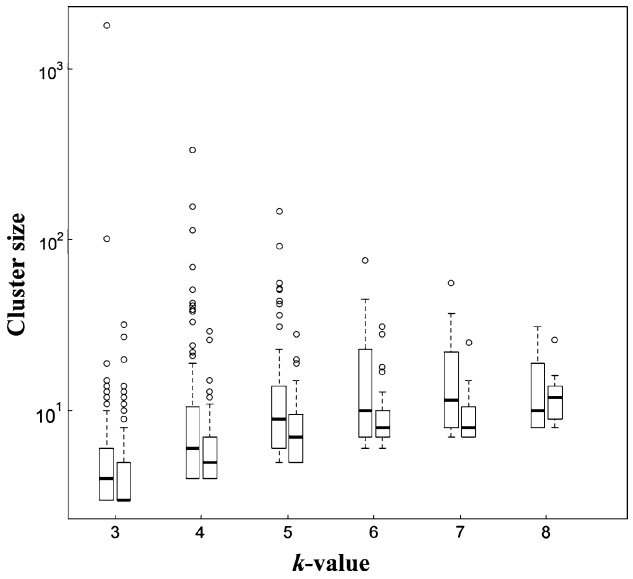

Fig. 4 Distribution of cluster sizes in the *A. thaliana* PPI network. The *horizontal axis* of this *box* plot denotes the *k*-value, and the *vertical axis* shows the cluster size (i.e., the sum of proteins in a cluster). For each *k*-value, clusters containing proteins that may interact with *R. solanacearum* are shown on the *left*, while the other clusters are plotted on the *right*. According to Wilcoxon rank-sum tests, the difference between these two kinds of clusters is significant ($p < 0.005$ for $k = 3$ and 4, and $p < 0.05$ for $k = 5$, 6 and 7). In each box, the *center band* is the median, while the *top* and *bottom lines* are the upper and lower quartiles, respectively. The ends of the whiskers are within a 1.5 inter-quartile range (IQR, which is equal to the difference between the upper quartile and the lower quartile). All outliers beyond the whiskers are shown as *circles*

interolog method. They reported that PPIs can be transferred when a pair of proteins has sequence identity >80% or *E*-value $<10^{-70}$. Although the thresholds in predicting intra-species PPIs should be different from those of inter-species PPI prediction, we realized the thresholds of the interolog method used in our work (0.001 BLAST *E*-value, 30% sequence identity and 80% aligned sequence length coverage) seem quite low. Furthermore, the domain-based method should be regarded as a special case of the interolog method, which focuses on the scale of domains. Therefore, the cut-offs of the domain-based method used in our work (i.e. 0.001 Pfam *E*-value and 90% aligned sequence length coverage) also seem too low. Due to the lack of experimentally determined PPIs between *R. solanacearum* and *A. thaliana*, we are not able to quantitatively estimate the reliability of the predicted PPIs based on the current thresholds. On the one hand, such low thresholds result in a relatively large PPI network, which provides sufficient data for network analysis. On the other hand, such thresholds also lead to a high false positive rate in the predicted PPIs.

To quantify the sizes of predicted PPIs based on different thresholds, we repeated the prediction using more stringent thresholds. Compared with the interolog method, the domain-based method generally shows lower accuracy (Rhodes et al. 2005). Therefore, we only investigated the interolog method's performance using a series of *E*-value and sequence identity. As expected, the number of predicted PPIs was dramatically decreased with more stringent thresholds used (Fig. 5). For instance, no PPI could be predicted when the sequence identity cut-off was set to 65% and only 170 PPIs, which covered 14 *R. solanacearum* proteins and 124 *A. thaliana* proteins, could be predicted when the *E-value* cut-off was assigned as 10^{-70} (Fig. 5). Interestingly, we still found that the clusters containing proteins that are predicted to interact with *R. solanacearum* proteins are always significantly larger than other clusters in *A. thaliana* PPI network, although the number of predicted PPIs is highly affected by choosing more stringent thresholds (Fig. 5). This suggests that the current prediction based on low thresholds is still helpful to explore some global characteristics of PPIs between *R. solanacearum* and *A. thaliana*.

Database server

We have made the predicted PPI data freely accessible at http://protein.cau.edu.cn/ppira. Users can query an *R. solanacearum* or an *A. thaliana* protein using different types of gene names. Potential interacting partners of the query protein in the opposite organism will be returned. To provide users with information about the biological function of the interacting partners and how they were predicted, essential prediction parameters are included such as *E*-value, sequence identity and aligned sequence length coverage. In addition, GO annotations of each potential interacting partner are listed. Due to the low thresholds used in the two PPI prediction methods, users may further estimate the reliability of a predicted PPI by checking the accompanied prediction parameters.

Conclusions

Using two well-known PPI prediction methods, we identified 3,074 potential PPIs between *R. solanacearum* and its plant host, *A. thaliana*. Due to internal limitations of the computational methods, the predicted data may still suffer from two drawbacks. First, the predicted PPI network is still far from complete. Second, the predicted data may inevitably contain a lot of false positives. To quantitatively assess the reliability of the predicted PPIs, experimentally determined PPI data are required. Even so, the predicted PPI data have allowed us to catch a glimpse of the overall picture of the PPI network between *R. solanacearum* and *A. thaliana*. We hope that the current work can shed light

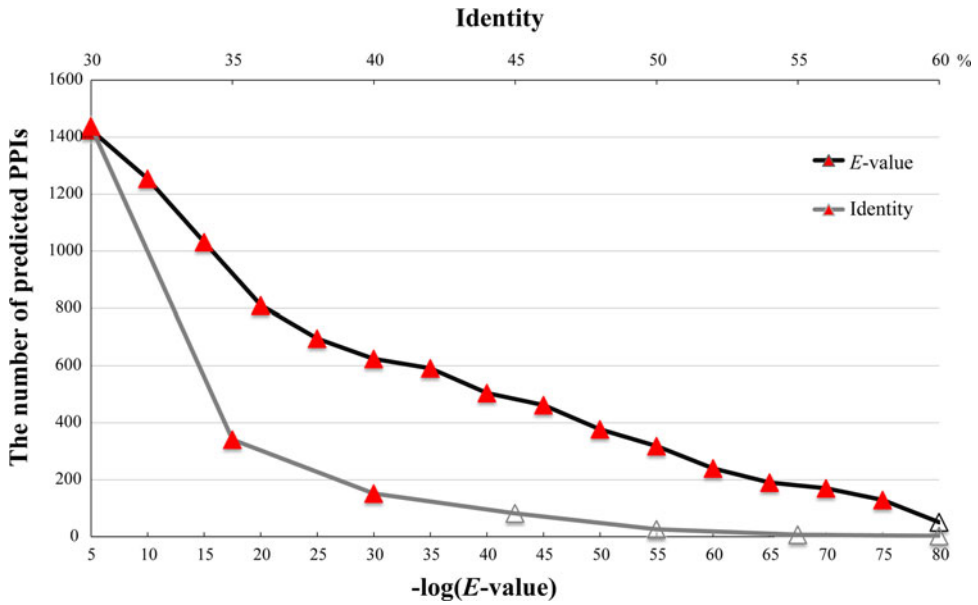

Fig. 5 The prediction results of the interolog method based on different *E*-value or sequence identity cut-offs. The *bottom and top horizontal axes* represent *E*-value and sequence identity, respectively. The *vertical axis* stands for the number of predicted PPIs between *R. solanacearum* and *A. thaliana*. The interolog method's performance on different thresholds of *E*-value (*black line*) and sequence identity (*gray line*) was evaluated, respectively. That is to say, we evaluated the performance by changing a series of cut-offs of one parameter and keeping the other one unchanged. Note that the aligned sequence length coverage cut-off was always fixed at 80% in the above computational experiments. The *solid triangle* denotes that clusters containing proteins that may interact with *R. solanacearum* are significantly larger in *A. thaliana* PPI network (Wilcoxon rank-sum test, $p < 0.05$; $k = 3$). The *hollow triangle* indicates that the number of the corresponding clusters was <10 and we did not perform statistical test

for further research into the molecular pathogenesis of *R. solanacearum*. For instance, the predicted data may inspire a path to the discovery of new anti-bacterial drug targets.

Acknowledgments We thank Ting-You Wang and Yuan Zhou at the bioinformatics center of China Agricultural University for helpful discussions. This research was supported by grants from the National Natural Science Foundation (30830058, 31070259 and J0730639) and the State Key Laboratory of Agrobiotechnology (2010SKLAB05-11).

References

Adamcsek B, Palla G, Farkas IJ, Derenyi I, Vicsek T (2006) Cfinder: locating cliques and overlapping modules in biological networks. Bioinformatics 22(8):1021–1023

Assenov Y, Ramirez F, Schelhorn SE, Lengauer T, Albrecht M (2008) Computing topological parameters of biological networks. Bioinformatics 24(2):282–284

Bateman A, Birney E, Durbin R, Eddy SR, Howe KL, Sonnhammer EL (2000) The pfam protein families database. Nucleic Acids Res 28(1):263–266

Bendtsen JD, Nielsen H, von Heijne G, Brunak S (2004) Improved prediction of signal peptides: Signalp 3.0. J Mol Biol 340(4): 783–795

Bishop JG, Dean AM, Mitchell-Olds T (2000) Rapid evolution in plant chitinases: molecular targets of selection in plant-pathogen coevolution. Proc Natl Acad Sci USA 97(10):5322–5327

Bogdanove AJ (2002) Protein-protein interactions in pathogen recognition by plants. Plant Mol Biol 50(6):981–989

Boudsocq M, Willmann MR, McCormack M, Lee H, Shan L, He P, Bush J, Cheng SH, Sheen J (2010) Differential innate immune signalling via ca(2+) sensor protein kinases. Nature 464(7287): 418–422

Cui J, Li P, Li G, Xu F, Zhao C, Li Y, Yang Z, Wang G, Yu Q, Shi T (2008) Atpid: *Arabidopsis thaliana* protein interactome database—an integrative platform for plant systems biology. Nucleic Acids Res 36(Database issue):D999–D1008

Derenyi I, Palla G, Vicsek T (2005) Clique percolation in random networks. Phys Rev Lett 94(16):160202

Du L, Ali GS, Simons KA, Hou J, Yang T, Reddy AS, Poovaiah BW (2009) Ca(2+)/calmodulin regulates salicylic-acid-mediated plant immunity. Nature 457(7233):1154–1158

Florez AF, Park D, Bhak J, Kim BC, Kuchinsky A, Morris JH, Espinosa J, Muskus C (2010) Protein network prediction and topological analysis in leishmania major as a tool for drug target selection. BMC Bioinform 11:484

Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talon M, Dopazo J, Conesa A (2008) High-throughput functional annotation and data mining with the blast2go suite. Nucleic Acids Res 36(10):3420–3435

Hayward AC (1991) Biology and epidemiology of bacterial wilt caused by *Pseudomonas solanacearum*. Annu Rev Phytopathol 29:65–87

He F, Zhang Y, Chen H, Zhang Z, Peng YL (2008) The prediction of protein–protein interaction networks in rice blast fungus. BMC Genomics 9:519

Iwakawa H, Shinmyo A, Sekine M (2006) Arabidopsis cdka;1, a cdc2 homologue, controls proliferation of generative cells in male gametogenesis. Plant J 45(5):819–831

Jonsson PF, Bates PA (2006) Global topological features of cancer proteins in the human interactome. Bioinformatics 22(18):2291–2297

Jonsson PF, Cavanna T, Zicha D, Bates PA (2006) Cluster analysis of networks generated through homology: automatic identification of important protein communities involved in cancer metastasis. BMC Bioinform 7:2

Kerrien S, Alam-Faruque Y, Aranda B, Bancarz I, Bridge A, Derow C, Dimmer E, Feuermann M, Friedrichsen A, Huntley R, Kohler C, Khadake J, Leroy C, Liban A, Lieftink C, Montecchi-Palazzi L, Orchard S, Risse J, Robbe K, Roechert B, Thorneycroft D, Zhang Y, Apweiler R, Hermjakob H (2007) Intact—open source resource for molecular interaction data. Nucleic Acids Res 35(Database issue):D561–D565

Kim JG, Park D, Kim BC, Cho SW, Kim YT, Park YJ, Cho HJ, Park H, Kim KB, Yoon KO, Park SJ, Lee BM, Bhak J (2008) Predicting the interactome of *Xanthomonas oryzae pathovar oryzae* for target selection and db service. BMC Bioinform 9:41

Krogh A, Larsson B, von Heijne G, Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. J Mol Biol 305(3):567–580

Matthews LR, Vaglio P, Reboul J, Ge H, Davis BP, Garrels J, Vincent S, Vidal M (2001) Identification of potential interaction networks using sequence-based searches for conserved protein–protein interactions or "interologs". Genome Res 11(12):2120–2126

Moller S, Croning MD, Apweiler R (2001) Evaluation of methods for the prediction of membrane spanning regions. Bioinformatics 17(7):646–653

Ng SK, Zhang Z, Tan SH (2003) Integrative approach for computationally inferring protein domain interactions. Bioinformatics 19(8):923–929

Niebel A, de Almeida Engler J, Hemerly A, Ferreira P, Inze D, Van Montagu M, Gheysen G (1996) Induction of cdc2a and cyc1at expression in *Arabidopsis thaliana* during early phases of nematode-induced feeding cell formation. Plant J 10(6):1037–1043

Ogmen U, Keskin O, Aytuna AS, Nussinov R, Gursoy A (2005) Prism: protein interactions by structural matching. Nucleic Acids Res 33(Web Server issue):W331–W336

Palla G, Derenyi I, Farkas I, Vicsek T (2005) Uncovering the overlapping community structure of complex networks in nature and society. Nature 435(7043):814–818

Pinzon A, Rodriguez RL, Gonzalez A, Bernal A, Restrepo S (2011) Targeted metabolic reconstruction: a novel approach for the characterization of plant-pathogen interactions. Brief Bioinform 12(2):151–162

Rhodes DR, Tomlins SA, Varambally S, Mahavisno V, Barrette T, Kalyana-Sundaram S, Ghosh D, Pandey A, Chinnaiyan AM (2005) Probabilistic model of the human protein–protein interaction network. Nat Biotechnol 23(8):951–959

Salanoubat M, Genin S, Artiguenave F, Gouzy J, Mangenot S, Arlat M, Billault A, Brottier P, Camus JC, Cattolico L, Chandler M, Choisne N, Claudel-Renard C, Cunnac S, Demange N, Gaspin C, Lavie M, Moisan A, Robert C, Saurin W, Schiex T, Siguier P, Thebault P, Whalen M, Wincker P, Levy M, Weissenbach J, Boucher CA (2002) Genome sequence of the plant pathogen *Ralstonia solanacearum*. Nature 415(6871):497–502

Salwinski L, Miller CS, Smith AJ, Pettit FK, Bowie JU, Eisenberg D (2004) The database of interacting proteins: 2004 update. Nucleic Acids Res 32(Database issue):D449–D451

Shoemaker BA, Panchenko AR (2007) Deciphering protein–protein interactions. Part ii. Computational methods to predict protein and domain interaction partners. PLoS Comput Biol 3(4):e43

Sonnhammer EL, von Heijne G, Krogh A (1998) A hidden markov model for predicting transmembrane helices in protein sequences. Proc Int Conf Intell Syst Mol Biol 6:175–182

Spok A, Stubenrauch G, Schorgendorfer K, Schwab H (1991) Molecular cloning and sequencing of a pectinesterase gene from *Pseudomonas solanacearum*. J Gen Microbiol 137(1):131–140

Stahl EA, Bishop JG (2000) Plant–pathogen arms races at the molecular level. Curr Opin Plant Biol 3(4):299–304

Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, Tyers M (2006) Biogrid: a general repository for interaction datasets. Nucleic Acids Res 34(Database issue):D535–D539

Stéphane G, Christian B (2002) *Ralstonia solanacearum*: secrets of a major pathogen unveiled by analysis of its genome. Mol Plant Pathol 3(3):111–118

Storey JD (2002) A direct approach to false discovery rates. J Royal Stat Soc: Series B (Stat Methodol) 64(3):479–498

Swarbreck D, Wilks C, Lamesch P, Berardini TZ, Garcia-Hernandez M, Foerster H, Li D, Meyer T, Muller R, Ploetz L, Radenbaugh A, Singh S, Swing V, Tissier C, Zhang P, Huala E (2008) The arabidopsis information resource (tair): gene structure and function annotation. Nucleic Acids Res 36(Database issue):D1009–D1014

Wu X, Zhu L, Guo J, Zhang DY, Lin K (2006) Prediction of yeast protein–protein interaction network: insights from the gene ontology and annotations. Nucleic Acids Res 34(7):2137–2150

Yu H, Luscombe NM, Lu HX, Zhu X, Xia Y, Han JD, Bertin N, Chung S, Vidal M, Gerstein M (2004) Annotation transfer between genomes: protein–protein interologs and protein-DNA regulogs. Genome Res 14(6):1107–1118

Yu H, Kim PM, Sprecher E, Trifonov V, Gerstein M (2007) The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. PLoS Comput Biol 3(4):e59