Bradyrhizobium diazoefficiens USDA 110-*Glycine max* interactome provides candidate proteins associated with symbiosis

Li Zhang[1,†], Jin-Yang Liu[2,†], Huan Gu[2], Yanfang Du[3], Jian-Fang Zuo[3], Zhibin Zhang[3], Menglin Zhang[3], Pan Li[5], Jim M. Dunwell[6], Yangrong Cao[7], Zuxin Zhang[4,*] and Yuan-Ming Zhang[1,*]

1 College of Plant Science and Technology, Huazhong Agricultural University, Wuhan 430070, China / Xinxiang Key Laboratory of Public Health Informatics, School of Public Health, Xinxiang Medical University, Xinxiang 453003, China

2 College of Agriculture, Nanjing Agricultural University, Nanjing 210095, China

3 College of Plant Science and Technology, Huazhong Agricultural University, Wuhan 430070, China

4 National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China

5 Xinxiang Key Laboratory of Public Health Informatics, School of Public Health, Xinxiang Medical University, Xinxiang 453003, China

6 School of Agriculture, Policy and Development, University of Reading, Reading RG6 6AR, United Kingdom

7 College of Life Science and Technology, Huazhong Agricultural University, Wuhan 430070, China

†: These authors contributed equally to this work.

* **Correspondences** (e-mails soyzhang@mail.hzau.edu.cn or zuxinzhang@mail.hzau.edu.cn)

College of Plant Science and Technology, Huazhong Agricultural University, Wuhan, China

Data Availability Statement All the datasets analyzed were from previously published datasets. Supporting Information may be found in additional files.

Funding This work was supported by the National Natural Science Foundation of China (31571268), Huazhong Agricultural University Scientific & Technological Self-innovation Foundation (Program No. 2014RC020) and State Key Laboratory of Cotton Biology Open Fund (CB2017B01). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests The authors have declared that no competing interests exist.

Abstract

Although the legume-rhizobium symbiosis is a most important biological process, there is a limited knowledge about the protein interaction network between host and symbiont. Using interolog and domain-based approaches, we constructed an inter-species protein interactome with 5115 protein-protein interactions between 2291 *Glycine max* and 290 *Bradyrhizobium diazoefficiens* USDA 110 proteins. The interactome was validated by expression pattern analysis in nodules, GO term semantic similarity, and co-expression analysis. One sub-network was further confirmed using luciferase complementation image assay. In the *G. max-B. diazoefficiens* interactome, bacterial proteins are mainly ion channel and transporters of carbohydrates and cations, while *G. max* proteins are mainly involved in the processes of metabolism, signal transduction, and transport. We also identified the top ten highly interacting proteins (hubs) for each of the two species. KEGG pathway analysis for each hub showed that two 14-3-3 proteins (SGF14g and SGF14k) and five heat shock proteins in *G. max* are possibly involved in symbiosis, and ten hubs in *B. diazoefficiens* may be important symbiotic effectors. Subnetwork analysis showed that 18 symbiosis-related SNARE proteins may play roles in regulating bacterial ion channels, and SGF14g and SGF14k possibly regulate the rhizobium dicarboxylate transport protein DctA. The predicted interactome and symbiosis proteins provide a valuable basis for understanding the molecular mechanism of root nodule symbiosis in soybean.

Keywords: root nodule symbiosis; interactome; nitrogen fixation; protein-protein interaction

Introduction

Rhizobia are gram-negative soil bacteria and have the ability to establish a nitrogen-fixing symbiosis on the roots of legume plants [1,2]. This legume-rhizobium symbiosis is of great agronomic importance and allows the plant to grow successfully in the absence of externally supplied nitrogen fertilizer [1]. Using the legume-rhizobium symbiosis to improve soil fertility is also an effective way to rehabilitate infertile land.

Among rhizobia, *Bradyrhizobium diazoefficiens* USDA 110 (previously named *Bradyrhizobium japonicum* USDA 110) is the most agriculturally important rhizobial bacterium as it is able to specifically infect soybean (*Glycine max*), one of the important legume plants in the world, and form a nitrogen-fixing symbiosis [3]. Furthermore, *G. max*-*B. diazoefficiens* is one of the most studied soybean-rhizobium symbiotic models [4]. Given the importance of such unique feature of legumes, further studies on the mechanisms of the soybean-rhizobium symbiosis are of particular interest. Importantly, the genome sequences of both *B. diazoefficiens* USDA 110 and *G. max* are now available [3,5], and provide an opportunity to better understand the mechanism of symbiotic features in terms of genomics and proteomics.

In *B. diazoefficiens* USDA 110, several genes related to various stages of the symbiosis process have been identified [3]. In soybean, comparative genomics analysis of legumes also predicted several nodulin genes [5]. Additionally, microarray approaches and RNA-seq analysis in soybean revealed a large number of genes differentially regulated during the symbiosis [4,6,7]. However, none of the above studies have focused on the complex interactions between candidate symbiosis-related genes. Generally, the proteins in the symbiosis process function as a complex network, which combines complex chemical, physical and biological interactions between rhizobial bacteria and their host plants [8]. To better elucidate the complex microbial

communities and investigate the mechanism of nitrogen-fixing symbiosis, it is necessary to construct the protein interactions between rhizobium and their host legume plants [9].

For any host-microbe system (including legume-rhizobium symbiosis and host-pathogen system), it is important to understand the mechanism by which the symbiotic or pathogenic bacteria can infect its host. As is known, one of the infection processes of any host-pathogen system is via protein-protein interactions (PPIs) between pathogen proteins and their host proteins [10]. PPIs are the associations of proteins with each other. They play crucial roles in the infection process and in initiating a defense response [11-13]. To date there have been several studies that have focused on the interactions among the protein networks of a host and a pathogen, and identified many new candidate proteins associated with the invasion [11,13-16]. However, PPI network analyses between two species have not been applied to legume-rhizobium symbiosis studies. Therefore, we attempted to construct the PPI interactome between soybean proteins and *B. diazoefficiens* USDA 110 proteins at a genome scale; such an investigation represents a critical step for studying the molecular basis of soybean-rhizobium symbiosis.

In the past decade, a series of computational approaches for PPI prediction have been developed [16,17], and these now play important roles in complementing the various experimental approaches. The existing computational approaches for PPI prediction have exploited diverse data features, which include domain and motif information [18-21], network topology [21,22], gene ontology (GO) [18-20], gene expression [18,19], protein sequence similarity [14,23], and pathway analysis [24]. At present, the interolog and domain-based approaches [25-27] are widely used [14,15,28]. The interolog method is based on protein sequence similarity to conduct the PPI prediction, which maps interactions in the source organism onto the target organism to find possible interactions in the target organism [25,26]. The domain-based method uses

domain interaction information and relies on the principle that if a protein pair contains an interacting domain pair, the two proteins are expected to interact with each other [27].

In this study, we predicted a protein-protein interaction network between *G. max* and *B. diazoefficiens* USDA 110 using both interolog and domain-based methods. GO annotation and gene expression data were utilized to validate the quality of the predicted PPI network. PANTHER overrepresentation test and KEGG pathway enrichment analysis were conducted to determine the biological function of the *B. diazoefficens* and *G. max* proteins predicted in the PPI network. We analyzed the subnetworks of the protein interactome to identify the candidate proteins possibly related to the soybean-rhizobium symbiosis, and used luciferase complementation image (LCI) assay [29,30] to confirm a subnetwork with two 14-3-3 proteins. In addition, we discuss how these predicted PPIs can help us to better understand this process.

Results

Network construction

Based on the well-studied experimental PPIs of seven model organisms: *Arabidopsis thaliana*, *Caenorhabditis elegans*, *Drosophila melanogaster*, *Escherichia coli* K12, *Homo sapiens*, *Mus musculus* and *Saccharomyces cerevisiae*, the PPIs between *G. max* and *B. diazoefficiens* were predicted in this study. To make use of more comprehensive information, we obtained the PPIs of seven organisms from multiple databases: BioGrid [31], DIP [32], HPRD [33], IntAct [34], MINT [31] and TAIR [35]. An ID dictionary was obtained from BioGrid to provide cross-database ID mapping. For mismatching IDs, we corrected manually in the Uniprot ID mapping server. As a result, we incorporated 44702 PPIs with 9948 proteins in *A. thaliana*, 28791 PPIs with 11543 proteins in *C. elegans*, 78383 PPIs with 9438 proteins in *D.*

melanogaster, 24460 PPIs with 3358 proteins in *E. coli* K12, 281387 PPIs with 15937 proteins in *H. sapiens*, 31010 PPIs with 8567 proteins in *M. musculus*, and 311333 PPIs with 6149 proteins in *S. cerevisiae* (Table S1).

All the 56044 *G. max* and 8317 *B. diazoefficiens* USDA 110 proteins were used to conduct a genome-wide PPI prediction. Among 8317 *B. diazoefficiens* proteins, 2356 proteins are secreted or membrane proteins (Table S2), which have the possibility to interact with *G. max* proteins. Using the pipeline shown in Figure 1 and filtered by above 2356 secreted or membrane proteins, 5115 PPIs between 2291 soybean proteins and 290 *B. diazoefficiens* USDA 110 proteins were predicted (Figure S1; Table S3). In addition, 233545 intra-species PPIs in soybean (Table S4) and 11106 intra-species PPIs in *B. diazoefficiens* USDA 110 (Table S5) were predicted. In summary, there were a total of 249766 PPIs, including inter- and intra-species PPIs, and 54471 PPIs (21.81%) were found in more than one species or experiment. All predicted interactions and the detailed annotation information of the proteins are available in Tables S3 to S5.

Quality assessment of protein–protein interactions

To date, few experimental PPIs between *B. diazoefficiens* USDA 110 and *G. max* have been identified, so it is difficult to validate the predicted PPI network by experimental approaches. For this reason, computational biology approaches were used to validate the quality of the predicted PPI network. In this study, we analyzed the gene expression pattern in nodules of all the soybean and rhizobium proteins in the *G. max*-*B. diazoefficiens* interactome. Furthermore, we conducted GO term semantic similarity [23,36] and co-expression analysis [28,37] of the intra-species PPI interactome. The results were used to deduce the quality of the *G. max*-*B. diazoefficiens* interactome, owing to the same methodologies.

Expression pattern in nodules The interaction between rhizobium and its host

legume results in the formation of a novel plant organ, the nodule. In nodules, the legume host interacts with rhizobium and exchanges photosynthetic products for ammonia from the rhizobial bacteria [38]. Thus, the predicted 5115 interactions between *B. diazoefficiens* and *G. max* are more likely to occur in nodules. In other words, most genes that encode the 2291 soybean proteins and the 290 *B. diazoefficiens* proteins in 5115 PPIs should be expressed in nodules. Analysis of the transcriptome data showed that 71.80% (1644) soybean genes were expressed in nodules with FPKM > 5. However, for the whole genome, the percent of genes expressed in nodules with FPKM > 5 is only 33.34% (18686 genes of the entire genome, which has 56045 genes). This indicates that most soybean genes in the above predicted network were indeed significantly expressed in nodules.

In previous studies, genome-wide analysis of *B. diazoefficiens* genes in symbiosis bacteroids was conducted at the transcriptome [39,40] and protein [41] levels. And these datasets were also used to investigate the expression patterns of 290 *B. diazoefficiens* genes in soybean root nodules. As a result, 172 (59.31%) genes were found to be expressed in symbiosis bacteroids (Table S6).

Functional similarity based on GO annotation Two interacting proteins would have similar or related functions and should share some common GO annotations [23,28,36]. Thus, GO annotation information of two interacting proteins was used to measure the accuracy of our prediction. Among 56044 soybean genes, 30023 (53.57%) genes were annotated with at least one GO term in any of the three GO categories (molecular function, biological process, and cellular component). Of all the 233545 soybean PPIs, 128862, 66369 and 26135 PPIs were annotated in the categories of molecular function, biological process, and cellular component (125086, 63581, 25007 non-self interactions), respectively (Table S4).

To measure the semantic similarity between GO terms and to evaluate the reliability of predicted PPIs, three functional similarity scores, sim_{JC}^{BP}, sim_{JC}^{MF} and sim_{JC}^{CC},

were calculated using non-self interactions in each GO category. Meanwhile, randomly selected protein pairs of the same size served as a control. As a result, significant differences for each of three sim_{JC} scores between predicted PPIs and randomly selected protein pairs were observed (Figure 2). All the proportions of score 1.0 in sim_{JC}^{BP}, sim_{JC}^{MF}, and sim_{JC}^{CC} were significantly higher in predicted soybean PPIs than those in randomly selected protein pairs, indicating that the predicted interaction network indeed preferentially connects functionally related proteins.

Co-expression of predicted soybean PPIs Levels of mRNA expression have some relationship with protein-protein interactions [42]. The interacting proteins tend to have correlated gene expression patterns, especially for subunits of the same protein complex [28,37,43]. Thus, we investigated the relationship of our predicted intra-species PPIs with mRNA expression levels in soybean. In this study, we used the transcriptome data from nine tissues of *G. max* to investigate expression correlation between two interacting proteins. The co-expression level of two interacting proteins was calculated by a widely used measure, the Pearson correlation coefficient (PCC) [44].

Among 233545 soybean intra-species PPIs (Table S3), 216097 PCC scores were successfully calculated. Among these scores, 23.84% (51524) protein interactions had a high PCC score ($r > 0.6$). In randomly selected protein pairs, however, the proportion was only 13.80%. This implies that the predicted interacting pairs have a significant co-relationship and the predicted PPI networks have high reliability. For conserved PPIs identified from more than one species or experiment, 34.72% had a high PCC score ($r > 0.6$), indicating a higher reliability. This is consistent with the conclusion that protein interactions detected by more than one high-throughput interaction assay are more accurate [36,45].

Conserved PPIs identified in more than two species

Common protein interactions predicted from multiple species can be considered as evolutionarily conserved interactions that have very high confidence [36]. In this study, we detected common protein interactions from more than two species. As a result, 60 conserved PPIs including 54 *G. max* proteins and 21 *B. diazoefficiens* proteins in *G. max-B. diazoefficiens* interactome were found (Figure S2). Among these 54 *G. max* proteins, more importantly, 49 proteins were expressed with FPKM > 5 in the underground tissues (root, root hair and nodule) and 24 proteins had high expression levels with FPKM > 100 in the underground tissues.

Function enrichment analysis of proteins in *G. max-B. diazoefficiens* interactome

To determine whether any biological function biases exist in the *B. diazoefficens* and *G. max* proteins in the predicted PPI network, we classified the proteins using the PANTHER overrepresentation test and conducted KEGG pathway enrichment analysis. The corresponding results with Bonferroni correction are listed in Tables 1 and 2, respectively.

***B. diazoefficens* USDA 110 proteins** In the predicted PPI network, *B. diazoefficens* proteins are mainly ion channel and transporters of carbohydrates and cations (Table 1). As the legume-rhizobium interaction involves the bacterial fixation of atmospheric nitrogen in exchange for plant-produced carbohydrates and all the essential nutrients required for bacterial metabolism [38,46,47], these transporters may provide the opportunities for rhizobial nodulation. KEGG pathway enrichment analysis further showed that bacterial proteins in *G. max-B. diazoefficiens* interactome were involved in pathways associated with symbiosis, such as protein export, peptidoglycan biosynthesis, ABC transporters and the bacterial secretion system (Tables 2 and S7), which are consistent with those in previous studies [48-51].

***G. max* proteins** Protein classification in soybean showed that proteins interacting with *B. diazoefficiens* were mainly involved in the processes of gene transcription and

translation, transport, metabolism, and signal transduction (Table 1). In transport, they were ion channels, ATP-binding cassette (ABC) transporters, mitochondrial carrier proteins and amino acid transporters. In signal transduction, 34 G-proteins, 18 small GTPase, 32 calmodulin and 18 SNARE proteins were present in the predicted PPIs and directly interacted with bacteria (Tables 1 and S8). Moreover, KEGG pathway enrichment analysis showed that soybean proteins in the predicted PPIs were involved in carbon metabolism, tricarboxylic acid cycle and N-glycan biosynthesis (Tables 2 and S7). Consistent with the above observations, Carvalho *et al*. [7] demonstrated that soybean genes involved in signal transduction, transcriptional regulation and primary metabolism were induced by the presence of the rhizobial bacteria. Additionally, by comparing with the *G. max* nodulation-related genes or searching for homologs of *M. truncatula* and *L. japonicus* nodulation-related genes in previous studies [4,5,52], we investigated whether some *G. max* proteins in predicted PPIs are experimentally nodulation-related genes. As a result, 9 soybean nodulation-related genes were identified and their PPIs are list in Table S9. These results suggest that soybean proteins interacting with the rhizobium were involved in various specific areas of metabolism, and the predicted interactions may provide useful information to understand the molecular mechanism of the legume-rhizobium symbiosis.

Hubs in *G. max*-*B. diazoefficiens* interactome

In protein–protein interaction networks, most proteins (nodes) connect with few proteins, whereas, a small percentage of proteins interact with a large number of other proteins [53,54]. Such proteins (nodes) with a large number of interactions are called hubs, and are more essential than proteins with only a small number of interactions. These proteins are known to perform vital roles in various cellular processes under a range of conditions including those caused by host-pathogen interactions [53-56]. In the present study, we listed the top ten hubs of each species in the *G. max*-*B. diazoefficiens* interactome (Table 3). To further understand the functions of the twenty

hubs, we performed KEGG pathway enrichment analysis for the proteins interacting with each of the twenty hubs. These results are listed in Supplementary Table S10.

In soybean, the top ten hubs included two 14-3-3 proteins, a Pumilio 7 protein, five heat shock proteins (HSPs) and two ADP/ATP carrier proteins (Table 3). The KEGG pathways for the two 14-3-3 proteins contained two-component systems (TCSs), Tryptophan metabolism and Oxidative phosphorylation (Table S10). Pumilio 7 protein and two ADP/ATP carrier proteins were both involved in the processes of Oxidative phosphorylation and Glycerophospholipid metabolism. Three of the five HSPs were enriched to show interaction with bacterial proteins in the metabolism of glycerophospholipids (Table S10), which are important components of membrane lipids in bacteria.

In *B. diazoefficiens*, the ten hubs included BAC49080, BAC52411, BAC49957, BAC52381, BAC45806, BAC45833, BAC47677, BAC47750, BAC45992 and BAC46205 (Table 3). KEGG pathway enrichment analysis showed that seven hubs were involved in carbohydrate metabolism, including N-Glycan biosynthesis, Pyruvate metabolism, Glycolysis and Citrate cycle (Table S10).

Subnetworks related to symbiosis

Based on an analysis of the PPI networks, we can better understand the web of interactions that takes place inside a cell. One method to better understand the entire network is to partition it into a series of subnetworks. In the present study, we selected two subnetworks that separately contain SNAREs and 14-3-3 proteins for further analysis to identify candidate proteins related to symbiosis (Figure 3).

SNARE proteins SNARE proteins are vital for signal transduction and membrane fusion in plants [57,58]. There is now growing evidence that these proteins play crucial roles in symbiosis in legume nodules, such as those in *L. japonicus* [58] and *M. truncatula* [59,60]. In the present study, 18 SNARE proteins in *G. max* were

involved in the predicted *G. max-B. diazoefficens* interactome and closely interacted with *B. diazoefficens* proteins (Figure 3A), suggesting the critical roles of SNAREs in soybean root nodule symbiosis (RNS). Meanwhile, soybean SNAREs interacted with each other in Figure 3A, which was consistent with the results in previous structural studies that SNAREs could form complexes by interacting with other SNAREs [57,58].

14-3-3 protein 14-3-3 proteins are abundant proteins in plants, and are involved in signaling pathways to regulate plant development and response to stimulus. Li and Dhaubhadel [61] identified 18 genes (SGF 14a-r) coding 14-3-3 proteins in the whole soybean genome. Previous studies revealed that two of them (SGF14c and SGF14l) play critical roles in RNS [62] and homologs of SGF14b in *L. japonicus* were located in the peribacteroid membrane [63]. In our study, we found another two 14-3-3 proteins, Glyma.14G176900 (SGF14k) and Glyma.02G208700 (SGF14g), which are hubs that interacted with *B. diazoefficiens* to a high degree (Table 3). More importantly, we found that SGF14k and SGF14g were connected with four soybean nodulation genes, *Glyma.06G065600* (Nodulin26) [64], *Glyma.17G13300* (WD40 protein; homologs of MtCCS52) [65], *Glyma.17G193800* (nucleoporin; homologs of LjNUP85) [66] and *Glyma.14G008200* (nucleoporin; homologs of LjNUP133) (Figure 3B) [67]. The results of the predicted PPIs of SGF14k and SGF14g demonstrated that SGF14k and SGF14g were involved in RNS.

Validation of a subnetwork containing two 14-3-3 proteins using luciferase complementation image (LCI) assay experiment

Luciferase complementation image (LCI) assay is a well-established method to verify the predicted PPIs in a laboratory setting. To validate the accuracy of the predicted interactions, a subnetwork in Figure 3B was selected to test the interactions *in vivo*. As a result, nine were confirmed (Figure 4). For example, SGF14k was interacted with BAC48988, similarly, SGF14k and BAC49563, SGF14k and Nodulin26,

SGF14g and BAC48988, SGF14g and BAC49563, SGF14g and Nodulin26, SGF14g and NUP85, Nodulin26 and BAC49735, and Glyma13G158600 and BAC49563 (Figures 3B and 4). Among the nine pairs of PPIs, six interacting protein pairs are produced between *G. max* and *B. diazoefficens*. More importantly, the interaction between Glyma.06G065600 (Nodulin) and Glyma.17G193800 (nucleoporin; homologs of LjNUP85) has been found to be involved in RNS [64,66]. Meanwhile, soybean SGF14g and SGF14k were both verified by LCI assay to interact with soybean Nodulin26 and *B. diazoefficiens* proteins BAC48988 and BAC49563, suggesting the critical roles of SGF14g and SGF14k in the establishment of RNS. Additionally, Nodulin26 was also found to be interacted with *B. diazoefficens* protein BAC49735. Taken together, the results demonstrated the reliability of our predicted PPIs, which can provide a useful guideline for future research.

Discussion

Network validation for the predicted PPIs in this study

The network predicted in this study is relatively reliable. The reasons are as follows. First, nine predicted PPIs in a sub-network containing two 14-3-3 proteins (SGF14g and SGF14k) showed an interaction signal via the LCI assay (Figure 4). Meanwhile, nine soybean nodulation-related genes predicted in this study have been experimentally confirmed to be involved in RNS (Table S9). Additionally, three computational biology approaches were used to validate the predicted network in this study. For example, significantly higher proportions of score 1.0 for the three simJC indicators in predicted soybean PPIs than those in randomly selected protein pairs indicates high quality of the *G. max-B. diazoefficiens* interactome (Figure 2); a significant higher proportion of predicted interaction pairs showed a co-relationship in their gene expression levels (PCC score > 0.6) than did randomly selected protein pairs; soybean genes expressed in nodules with FPKM > 5 had a significantly higher

proportion (71.80%) in the predicted network than those (33.34%) in the entire genome, while 59.31% *B. diazoefficiens* genes were found to be expressed in symbiosis bacteroids.

Soybean proteins in the predicted PPIs were involved with pathways associated with symbiosis

The infection transcriptome analysis confirmed that proteins involved in various areas of metabolism were triggered in the host plant by the presence of nitrogen-fixing bacteria [7,68]. In the process of transport, Sugiyama *et al*. [69] revealed that the soybean ABC transporters play important roles in legume-rhizobium symbiosis, and Clarke *et al*. [70] found by proteome analysis that transporters of sulfate, nitrate, peptides, and various metal ions like calcium, potassium and zinc are present on the soybean symbiosome membrane. Consistently, soybean ABC transporters and ion channels were predicted to interact with *B. diazoefficiens* proteins in the present study (Table 1). Since these transporters can facilitate the movement of nutrients between the symbionts and ensure the establishment of symbiosis, the candidate transport proteins in *G. max-B. diazoefficens* interactome can help our understanding of the role of transporters on the symbiosome membrane. In carbohydrate metabolism, soybean proteins involved in carbon metabolism, tricarboxylic acid cycle and N-glycan biosynthesis directly interacted with bacteria (Tables 2 and S7). Consistently, Libault *et al*. [68] and Carvalho *et al*. [7] showed that carbohydrate metabolism like the tricarboxylic acid cycle and glycolysis were induced by the presence of rhizobia in both roots and root hairs. These metabolic effects ensure the development of nodules by providing the carbon [71], while the host plant provides rhizobia with all the essential nutrients such as carbon required for bacterial metabolism [38]. Various signal transduction pathways play important roles in various stages of the symbiosis. They can coordinate the development of epidermal and cortical cells to ensure rhizobial invasion and nodule initiation [7,72]. Previous studies have confirmed the

involvement of many nod factors in the signal transduction processes such as G-protein coupled receptor signaling pathways [73,74], small GTPase mediated signal transduction [75,76], calmodulin [77], Soluble N-Ethylmaleimide Sensitive Factor Attachment Protein Receptor (SNARE) proteins [58,78] and the MAPK (Mitogen-activated protein kinase) cascade [79]. In the present study, 34 G-proteins, 18 small GTPase, 32 calmodulin and 18 SNARE proteins were present in the predicted PPIs and directly interacted with bacteria (Tables 1 and S8). The subnetworks of related signaling transduction provide opportunities to reveal whether and how these networks are interconnected, and then give insights into the mechanism of symbiosis.

Hubs in the predicted network played roles in symbiosis

In previous studies, HSPs were reported to be involved in the host-pathogen interaction [80] and to be induced during symbiosis in response to pathogens [81-83], suggesting that HSPs play critical roles in the response of plant cells to biotic stressors. HSPs have also been identified in the symbiosome membrane of soybean [84], *L. japonica* [63] and *M. truncatula* [85] by proteome analysis. Moreover, Brechenmacher *et al*. [6] reported that HSPs were up-regulated in soybean roots during the interaction between *G. max* and *Bradyrhizobium japonicum*. In the present study, five of the top ten soybean hubs interacting with *B. diazoefficens* are HSPs, and three hub HSPs interacted with *B. diazoefficens* proteins in the metabolism of glycerophospholipids, an important component of bacteria membrane lipids (Table S8). The results in this study give us insights that the five HSPs interacting with bacteria in the predicted PPIs are key players in the establishment of RNS.

The other two highly interacting hubs were SGF14k and SGF14g, which were shown to interact with *B. diazoefficens* proteins in the pathways of two-component systems (TCSs) and tryptophan metabolism (Table S10). TCSs are abundant signaling pathways in prokaryotes [86,87]. They could transduce extracellular signals into the

449 cell and regulate multiple cellular processes in response to environmental stimuli
450 [88,89]. More importantly, transcriptional regulators of TCS showed increased
451 expression in bacteroids during RNS [39]. For Tryptophan metabolism, Hunter [90]
452 showed that *Bradyrhizobia* with altered tryptophan metabolism frequently have
453 altered symbiotic properties, and changes in the level of indole-3-acetic acid (a
454 tryptophan metabolism product) that is involved in bacteria-plant interactions
455 [6,91,92]. Notably, Radwan and Wu [62] revealed that two homologs of the above
456 14-3-3 proteins play critical roles in RNS. In the present study, subnetwork analysis
457 showed that SGF14k and SGF14g interacted with four soybean nodulin genes
458 (*Glyma.06G065600*, *Glyma.17G13300*, *Glyma.17G193800* and *Glyma.14G008200*).
459 Among the four nodulin genes, two (*Glyma.06G065600* and *Glyma.17G193800*) were
460 verified to interact with SGF14k and SGF14g by LCI assay experiments (Figures 3B
461 and 4). Therefore, we deduce that *Glyma.14G176900* (SGF14k) and
462 *Glyma.02G208700* (SGF14g) are involved in the process of nodulation.
463
464 Carbon metabolism was found to be closely related to RNS [7,68]. Delmotte et al. [41]
465 identified several proteins involved in carbon metabolism in symbiosome membrane
466 of soybean, including a complete set of tricarboxylic acid cycle enzymes,
467 gluconeogenesis and pentose phosphate pathway enzymes, by integrated proteomic
468 and transcriptomic analysis. In the present study, seven hubs (BAC49080, BAC52411,
469 BAC45833, BAC47677, BAC47750, BAC45992 and BAC46205) were involved in
470 carbon metabolism, including N-Glycan biosynthesis, Pyruvate metabolism,
471 Glycolysis and Citrate cycle (Table S10). Additionally, enriched KEGG pathways
472 contained protein processing in endoplasmic reticulum, Glycosylphosphatidylinositol
473 (GPI)-anchor biosynthesis, Pentose phosphate pathway and Proteasome (Table S10).
474 Yuan *et al*. [4] found that genes involved in protein processing in endoplasmic
475 reticulum were differentially expressed between different developmental periods of
476 the soybean nodule. Roux *et al*. [93] revealed that genes involved in GPI-anchor
477 biosynthesis and proteasome function were found to be preferentially expressed in

plant nodules. Therefore, these ten hubs of *B. diazoefficens* may be important symbiotic effectors and play roles in symbiosis.

Subnetwork analysis provide insight into the mechanism of root nodule symbiosis

Subnetwork analysis of SNAREs showed that SNAREs mainly interacted with membrane transporters or related proteins (Figure 3A). In detail, Glyma.10G008300 interacted with a cation efflux system protein (BAC50315), two ABC transporter permease proteins (BAC51159 and BAC49765), a cation-transporting ATPase (BAC52318) and an ammonium transporter (BAC45878). Five SNARE proteins (Glyma.04G072700, Glyma.10G149000, Glyma.07G042400, Glyma.03G029700 and Glyma.01G137300) interacted with BAC49080, a cation-transporting ATPase. Glyma.10G149000 and Glyma.13G307600 interacted with a Na^+/H^+ exchanger (BAC46205). Sokolovski *et al*. [94] proved that a plasma membrane SNARE protein in *Nicotiana benthamiana* guard cells could regulate Ca^{2+} channels and also possibly target other ion channels. The results indicated that SNAREs in the symbiosome membrane may play roles in regulating bacteria ion channels. Further analysis of the role of SNARE proteins will provide novel insights into RNS.

Through the subnetwork analysis, two 14-3-3 proteins, SGF14k and SGF14g, not only interacted with soybean nodulins but also were closely connected with two bacterial DctA proteins, BAC48988 and BAC49563 (Figure 3B). DctA was an important transporter for C4-dicarboxylic acids, which are the main form of carbon and energy sources from host plant to *rhizobium* [95]. Notably, DctA was reported to be essential for symbiotic nitrogen fixation in *Sinorhizobium meliloti*, as well as other rhizobia [78,96]. The relationships between the above two 14-3-3 proteins and DctA proteins were further verified by LCI assay experiments (Figure 4). Taken together, the results indicated that 14-3-3 proteins SGF14g and SGF14k regulate rhizobium DctA.

Of course, the predicted results are still far from complete and may inevitably contain a lot of false positives, as the coverage and accuracy of predicted PPIs largely depend on the quality of interaction data sets and the ability to identify the orthologs from the model organisms. Even so, the predicted PPI networks have allowed us to have an insight into the overall picture of the PPI network between *G. max* and *B. diazoefficiens* USDA 110, which provide useful information to understand the molecular mechanism of the legume-rhizobium symbiosis.

Materials and methods

Datasets

A collection of 8317 protein sequences of *B. diazoefficiens* USDA 110 were downloaded from the Ensembl genomes database (ftp://ftp.ensemblgenomes.org/pub/bacteria/release-30/fasta/bacteria_0_collection/bradyrhizobium_diazoefficiens_usda_110/pep/) [97]. Soybean whole genome sequences (*G. max Wm82.a2.v1*) were obtained from Phytozome V10.3 (http://genome.jgi.doe.gov/pages/dynamicOrganismDownload.jsf?organism=PhytozomeV10) [98]. For genes with multiple transcripts, the longest protein sequence was chosen [99]. As a result, 56044 protein sequences were obtained in *G. max*.

To conduct the interolog analysis, we utilized the PPI information of seven well-studied model organisms, namely *Arabidopsis thaliana*, *Caenorhabditis elegans*, *Drosophila melanogaster*, *Escherichia coli* K12, *Homo sapiens*, *Mus musculus* and *Saccharomyces cerevisiae*. Experimentally verified PPIs of the aforementioned seven organisms were obtained from the public protein-protein interaction databases: BioGrid, DIP, HPRD, IntAct, MINT and TAIR (Table S1). The ortholog information between the aforementioned seven organisms and *G. max* or *B. diazoefficiens* independently were obtained from InParanoid 8 [100].

To carry out the domain-based PPI prediction, we downloaded the interacting Pfam domain pairs from the database of protein domain interactions (DOMINE Version 2.0) [101], which contains a total of 26219 domain-domain interactions (DDI). To increase the accuracy of prediction, only 2989 high-confident domain pairs were used as reference in this study.

PPI prediction

Our PPI prediction was mainly based on the interolog method, along with the domain-based method to improve prediction accuracy. In the interolog method proposed by Walhout *et al*. [25], the pair of interactions A–B and A1–B1 are called an interolog if interacting proteins A and B in a species have interacting orthologs A1 and B1 in another species. Based on this theory, interolog PPI prediction is a process that maps interactions in the source organism onto the target organism to find possible interactions [26]. In the domain-based method, the two proteins are expected to interact with each other if a protein pair contains at least one interacting domain pair [27]. The protein domain annotations for *B. diazoefficiens* USDA 110 were conducted in the Pfam website [102], and the annotations for *G. max* proteins were obtained from Phytozome V10.3.

In this study, ortholog pairs between each of the aforementioned seven model organisms and *G. max* (or *B. diazoefficiens*) were obtained from the InParanoid database [100]. InParanoid scores between 0 and 1 reflect the relative evolutionary distance between orthologous gene pairs [103,104]. The top score 1.0 is the best blast hit and has high credibility, and orthologs with scores below 1.0 are more or less sensitive. To restrict the sensitivity, ortholog pairs were selected with a score cutoff of 0.5. These ortholog pairs were further divided into two groups according to InParanoid score: orthologs with top score 1.0 and ones with a score between 0.5 and 1.0. Using the interolog method, all the above ortholog pairs were mapped onto the integrated PPI interactomes of the seven model organisms to predict PPIs. The

predicted PPIs with low confidence orthologs were further filtered by the domain-based method to increase prediction accuracy and to decrease false positives (Figure 1).

Identification of secreted and membrane proteins in *B. diazoefficiens* USDA 110

The transmembrane and secreted proteins in *B. diazoefficiens* USDA 110 are considered to be positive candidates for interactions with *G. max*. All the proteins in *B. diazoefficiens* USDA 110 were used to predict transmembrane proteins through TMHMM 2.0 [105] and to identify secretory proteins through SingleIP 4.0 [106]. In TMHMM 2.0, 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. SingleIP 4.0 was employed with the default settings.

GO annotation and measurement of functional similarity

The GO annotations of *B. diazoefficiens* USDA 110 and *G. max* were obtained from the Gene Ontology Annotation (UniProt-GOA) Database [107] and Phytozome V10.3 [98], respectively. Semantic similarity scores between GO terms were measured by Jiang and Conrath's distance method and calculated in database FunSimMat [108,109] to evaluate the reliability of the predicted PPIs [110]. Jiang and Conrath's distance between two GO terms is based on information content and was defined as follows [111]:

$$sim_{JC}(t_1,t_2)=\frac{1}{IC(t_1)+IC(t_2)-2\times IC(MIA)+1}$$

$sim_{JC}(t_1,t_2)$ is the set of common ancestors of terms t_1 and t_2 in the ontology, and ranges between 0, for no similarity, to 1, for highest similarity. We used sim_{JC} for referring to this score. As GO annotation classifies functions of a protein according to three features: molecular function, biological process and cellular component, there

were, correspondingly, three independent sim_{JC} scores: sim_{JC}^{MF}, sim_{JC}^{BP} and sim_{JC}^{CC}.

Co-expression analysis

Transcriptome data of soybean were obtained from Phytozome V10.3 [98], which includes nine tissues (root, root hairs, nodules, leaves, stem, flower, pod, sam, and seed). The expression correlation between two interacting proteins was calculated using a widely used measure, Pearson correlation coefficient (PCC) [45]. The PCC value for each pair of non-self-interacting proteins was calculated using the Fragments Per Kilobase of transcript per Million mapped reads (FPKM) value of mRNA in the above nine tissues.

Luciferase Complementation Image (LCI) assays for PPIs in *Nicotiana benthamiana* cells

Materials Soybean (*G. max* Willimas 82) and tobacco plants were grown at 16-hlight / 8-h dark at 25℃ for 30-60 d. *B. japonicum* (USDA110) was grown on (HM) medium plates at containing 50 μg of chloramphenicol/ml for selection of plasmid 25℃.

RNA and DNA Isolation Soybean total RNA was isolated using the Trizol reagent (Invitrogen, Foster city, CA, USA) according to the manufacturer's instructions and the RNAs were treated with the DNase I (Promega). The first-strand cDNA was then synthesized using M-MLV reverse transcriptase (Promega). The total DNAs of the *Bradyrhizobium japonicum* was isolated according to the method of Casse et al. [112].

Primers and conditions for PCR Primers were analyzed by Oligo 6 (Table S11). PCR was carried out using a PCR system for 35 cycles (30 s at 95℃, 30 s at Tm and 1-4 mins at 72℃).

Luciferase Complementation Image (LCI) assays Full length coding sequence

of target genes were amplified by polymerase chain reaction from total RNA (Table S11) and were cloned into the *Bam*HI and *Sal*I sites of JW-771-N (NLUC), as well as *Kpn*I and *Sal*I sites of JW-772-C, to produce target gene-NLUC and target gene-CLUC recombination vectors for the LCI assay (for split Luc N-terminal/C-terminal fragment expression), respectively. Thus, N-gene, C-gene, N-LUC, and C-LUC were constructed according to previously described protocols. These constructs were transformed into Agrobacterium tumefaciens GV3101 strain through $CaCl_2$ transformation [113]. The p19 protein (tomato bushy stunt virus) was used to suppress gene silencing [114].

Detection of interactions in vivo The recombinant plasmids were transfected into *Agrobacterium tumefaciens* (GV3101). The OD600 of co-infiltrated *A. tumefaciens* strains is about 1.0 (gene-NLUC): 1.0 (gene-CLUC): 1.0 (P19), 500 μl of each, to co-culture for 2 h. Equal amount of the Agrobacterium suspension of each construct was mixed into a new 1.5 mL tube and vortexed for 10 sec to be ready for use. 8-10 weeks-old (16 h-light and 8 h-dark) *Nicotiana benthamiana* leaves were used to inject *A. tumefaciens* cocultures described above. Placed the tip end of the syringe (without needle) against the underside of the leaf (avoiding the veins) by supporting with one finger on the upperside, then gently pressed the syringe to infltrate the Agrobacterium mixture into the fresh leaf [115]. After growing for 48 h under the condition of 16 h-light and 8 h-dark, pieces of leaf abaxial epidermis were treated with 1 mM luciferin (promega, E1602), and the resulting luciferase signals were captured by Tanon-5200 image system (Tanon, Shanghai, China). To test each interacting protein pair, three experiments were performed and similar results were obtained.

Supporting information

S1 Table. Experimental protein-protein interactions of seven model species from public

databases

S2 Table. The selected 2,356 membrane and secreted proteins in *B. diazoefficiens* USDA 110

S3 Table. The predicted *G. max*-*B. diazoefficiens* interactome and detailed annotation information of the proteins, including 5115 inter-species PPIs between 2291 *G. max* and 290 *B. diazoefficiens* USDA 110 proteins

S4 Table. The predicted *G. max* interactome, including 233545 intra-species PPIs in soybean

S5 Table. The predicted *B. diazoefficiens* USDA 110 interactome, including 11106 intra-species PPIs in *B. diazoefficiens* USDA 110

S6 Table. List of 172 genes in the predicted PPIs that were detected to be expressed in bacteroids of the root nodule during symbiosis in at least one of three previous studies

S7 Table. List of input genes enriched in KEGG pathway enrichment analysis in Table 2 and their detailed annotation

S8 Table. Soybean proteins in the PPI network that were involved in signal transduction

S9 Table. Nodulation-related genes that experimentally interacted with *B. diazoefficiens* USDA 110 proteins

S10 Table. Top ten hubs of *G. max* and *B. diazoefficiens* USDA 110 in the *G. max*-*B. diazoefficiens* interactome and KEGG pathway enrichment analysis of these hubs by using their interacted proteins in the PPI interactome

S11 Table. Primers used in the luciferase complementation image (LCI) assays for PPIs in *Nicotiana benthamiana* cells

S1 Figure. Visualization of the predicted PPI network between soybean and *B. diazoefficiens* USDA 110. Each node represents a protein and each edge denotes an interaction. Red color circles represent soybean and yellow represent *B. diazoefficiens* USDA 110.

692 **S2 Figure. Conserved PPIs identified in more than two species**. Line represents the interaction
693 relationship, circle represents proteins; yellow circles are *B. diazoefficiens* USDA 110 proteins,
694 red, pink and grey circles are soybean proteins and respectively represent the expression values
695 FPKM > 100, 5 < FPKM ≤ 100 and FPKM < 5 in nodules.

696 Author contributions

697

698 YMZ conceived and designed the experiments, and revised the manuscript. ZXZ and
699 PL assisted the supervision of the LCI experiment and bioinformatics analysis,
700 respectively. LZ, HQ, ZBZ and MLZ performed bioinformatics analysis. JYL, YFD,
701 JFZ performed the LCI experiments. YRC provided materials and modified the
702 manuscript. LZ wrote the manuscript. All authors reviewed the manuscript.
703

704 References

705
706 1. Downie JA: Legume nodulation. Current biology 2014, **24**(5): R184-R190.
707 2. Peix A, Ramírez-Bahena MH, Velázquez E, Bedmar EJ: Bacterial associations with legumes. Crit. Rev.
708 Plant Sc. 2014, **34**: 17-42.
709 3. Kaneko T, Nakamura Y, Sato S, Minamisawa K, Uchiumi T, Sasamoto S, Watanabe A, Idesawa K, Iriguchi
710 M, Kawashima K, et al.: Complete genomic sequence of nitrogen-fixing symbiotic bacterium
711 *Bradyrhizobium japonicum* USDA110. DNA Research 2002, **9:** 189- 197.
712 4. Yuan S, Li R, Chen S, Chen H, Zhang C, Chen L, Hao Q, Shan Z, Yang Z, Qiu D, et al.: RNA-seq analysis
713 of differential gene expression responding to different rhizobium strains in soybean (*Glycine max*) roots.
714 *Front Plant Sci*. 2016, **7**: 721
715 5. Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, et al.:
716 Genome sequence of the palaeopolyploid soybean. Nature 2010, **463**: 178-183.
717 6. Brechenmacher L, Kim MY, Benitez M, Li M, Joshi T, Calla B, Lee MP, Libault M, Vodkin LO, Xu D, et
718 al.: Transcription profiling of soybean nodulation by *Bradyrhizobium japonicum*. *Mol. Plant Microbe*
719 *Interact.* 2008, **21**: 631-645.
720 7. Carvalho GA, Batista JS, Marcelino-Guimarães FC, Nascimento LC, Hungria M: Transcriptional analysis of
721 genes involved in nodulation in soybean roots inoculated with *Bradyrhizobium japonicum* strain CPAC 15.
722 BMC Genomics 2013, **14**: 153.
723 8. Remigi P, Zhu J, Young JP, Masson-Boivin C: Symbiosis within symbiosis: evolving nitrogen-fixing legume
724 symbionts. Trends Microbiol. 2016, **24**: 63-75.

9. Afroz A, Zahur M, Zeeshan N, Komatsu S: Plant-bacterium interactions analyzed by proteomics. *Front Plant Sci*. 2013, **4:** 21.

10. Qi Y, Noble WS: *Protein interaction networks: protein domain interaction and protein function prediction*. Springer Berlin Heidelberg, 2011.

11. Dyer MD, Neff C, Dufford M, Rivera CG, Shattuck D, Bassaganya-Riera J, Murali TM, Sobral BW: The human-bacterial pathogen protein interaction networks of Bacillus anthracis, Francisella tularensis, and Yersinia pestis. PLoS ONE 2010, **5**: e12089

12. Durmus Tekir SD, Ülgen KÖ: Systems biology of pathogen-host interaction: networks of protein-protein interaction within pathogens and pathogen-human interactions in the post-genomic era. Biotechnol. J. 2013, **8**: 85-96.

13. Martinez F, Rodrigo G, Aragones V, Ruiz M, Lodewijk I, Fernandez U, Elena SF, Daros JA: Interaction network of tobacco etch potyvirus NIa protein with the host proteome during infection. BMC Genomics 2016, **17:** 87.

14. Li ZG, He F, Zhang Z, Peng YL: Prediction of protein-protein interactions between *Ralstonia solanacearum* and *Arabidopsis thaliana*. Amino Acids 2012, **42:** 2363-2371.

15. Sahu SS, Weirick T, Kaundal R: Predicting genome-scale *Arabidopsis- Pseudomonas syringae* interactome using domain and interolog-based approaches. BMC Bioinformatics 2014, **15**: S13.

16. Nourani E, Khunjush F, Durmus S: Computational approaches for prediction of pathogen-host protein-protein interactions. Front Microbiol. 2015, **6:** 94.

17. Shen J, Zhang J, Luo X, Zhu W, Yu K, Chen K, Li Y, Jiang H: Predicting protein-protein interactions based only on sequences information. Proc. Natl. Acad. Sci. U S A. 2007, **104**: 4337- 4341.

18. Kshirsagar M, Carbonell J, Klein-Seetharaman J: Techniques to cope with missing data in host-pathogen protein interaction prediction. Bioinformatics 2012, **28:** i466-i472.

19. Kshirsagar M, Carbonell J, Klein-Seetharaman J: Multitask learning for host–pathogen protein interactions. Bioinformatics 2013, **29:** 217-226.

20. Coelho ED, Arrais JP, Matos S, Pereira C, Rosa N, Correia MJ, Barros M, Oliveira JL: Computational prediction of the human-microbial oral interactome. BMC Systems Biol. 2014, **8**: 1-12.

21. Tastan O, Yanjun QI, Carbonell JG, Kleinseetharaman J: Prediction of interactions between HIV-1 and human proteins by information integration. Pacific Symposium on Biocomputing Pacific Symposium on Biocomputing 2015, **527**: 516-527.

22. Qi Y, Tastan O, Carbonell JG, Klein-Seetharaman J, Weston J: Semi-supervised multi-task learning for predicting interactions between HIV-1 and human proteins. Bioinformatics 2011, **26**: i645- i652.

23. Gu H, Zhu P, Jiao Y, Meng Y, Chen M: PRIN: a predicted rice interactome network. BMC Bioinformatics 2011, **12**: 161.

24. Wuchty S: Computational prediction of host-parasite protein interactions between *P. falciparum* and *H. sapiens*. PLoS ONE 2011, **6**: e26960.

25. Walhout AJM, Sordella R, Lu XW, Hartley JL, Temple GF, Brasch MA, Thierry-Mieg N, Vidal M: Protein interaction mapping in *C. elegans* using proteins involved in vulval development. Science 2000, **287**:

116-122.

26. Matthews LR, Vaglio P, Reboul J, Ge H, Davis BP, 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.

27. Flórez AF, Park D, Bhak J, Kim BC, Kuchinsky A, Morris JH, Espinosa J, Muskus C: Protein network prediction and topological analysis in *Leishmania major* as a tool for drug target selection. BMC Bioinformatics 2010, **11**: 484.

28. Lei D, Lin R, Yin C, Li P, Zheng A: Global protein–protein interaction network of rice sheath blight pathogen. J. Proteome Res. 2014, **13:** 3277-3293.

29. Kerppola TK: Complementary methods for studies of protein interactions in living cells. Nature Methods 2006, **3**(12), 969-971.

30. Kerppola TK: Visualization of molecular interactions by fluorescence complementation. Nat Rev Mol. Cell Biol. 2006, **7**(6): 449-456.

31. Chatr-Aryamontri A, Breitkreutz BJ, Oughtred R, Boucher L, Heinicke S, Chen D, Stark C, Breitkreutz A, Kolas N, O'Donnell L, et al.: The BioGRID interaction database: 2015 update. Nucleic Acids Res. 2015, **43**, D470-D478.

32. Xenarios I, Salwinski L, Duan XJ, Higney P, Kim SM, Eisenberg D: DIP, the database of interacting proteins: a research tool for studying cellular networks of protein interactions. Nucleic Acids Res. 2002, **30**: 303-305.

33. Prasad TSK, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, Telikicherla D, Raju R, Shafreen B, Venugopal A, et al.: Human protein reference database -- 2009 update. Nucleic Acids Res. 2009, **37**: D767-D772.

34. Kerrien S, Aranda B, Breuza L, Bridge A, Broackes-Carter F, Chen C, Duesbury M, Dumousseau M, Feuermann M, Hinz U, et al.: The IntAct molecular interaction database in 2012. Nucleic Acids Res. 2012, **40**: D841-D846.

35. Rhee SY, Beavis W, Berardini TZ, Chen G, Dixon D, Doyle A, Garcia-Hernandez M, Huala E, Lander G, Montoya M, et al.: The *Arabidopsis* information resource (TAIR): a model organism database providing a centralized, curated gateway to *Arabidopsis* biology, research materials and community. Nucleic Acids Res. 2003, **31**: 224-228.

36. Lehner B, Fraser AG: A first-draft human protein-interaction map. Genome Biol. 2004, **5:** R63.

37. Jansen R, Greenbaum D, Gerstein M: Relating whole-genome expression data with protein- protein interactions. Genome Res. 2002, **12**: 37-46.

38. Udvardi M, Poole PS: Transport and metabolism in legume-rhizobia symbioses. Annu. Rev. Plant Biol. 2013, **64**: 781-805.

39. Pessi G, Ahrens CH, Rehrauer H, Lindemann A, Hauser F, Fischer HM, Hennecke H: Genome-wide transcript analysis of *Bradyrhizobium japonicum* bacteroids in soybean root nodules. Mol. Plant Microbe Interact. 2007, **20**: 1353-1363.

40. Cuklina J, Hahn J, Imakaev M, Omasits U, Forstner KU, Ljubimov N, Goebel M, Pessi G, Fischer HM,

Ahrens CH, et al.: Genome-wide transcription start site mapping of Bradyrhizobium japonicum grown free-living or in symbiosis - a rich resource to identify new transcripts, proteins and to study gene regulation. BMC Genomics 2016, **17**: 302.

41. Delmotte N, Ahrens CH, Knief C, Qeli E, Koch M, Fischer HM, Vorholt JA, Hennecke H, Pessi G: An integrated proteomics and transcriptomics reference data set provides new insights into the *Bradyrhizobium japonicum* bacteroid metabolism in soybean root nodules. Proteomics 2010, **10**: 1391-1400.
42. Grigoriev A: A relationship between gene expression and protein interactions on the proteome scale: analysis of the bacteriophage T7 and the yeast *Saccharomyces cerevisiae*. Nucleic Acids Res. 2001, **29**, 3513-3519.
43. Ge H, Liu ZH, Church GM, Vidal M: Correlation between transcriptome and interactome mapping data from *Saccharomyces cerevisiae*. Nat. Genet. 2001, **29**: 482-486.
44. Obayashi T, Kinoshita K: Rank of correlation coefficient as a comparable measure for biological significance of gene coexpression. DNA Research 2009, **16:** 249-260.
45. von Mering C, Krause R, Snel B, Cornell M, Oliver SG, Fields S, Bork P: Comparative assessment of large-scale data sets of protein-protein interactions. Nature 2002, **417**: 399-403.
46. Prell J, Poole P: Metabolic changes of rhizobia in legume nodules. Trends Microbiol. 2006, **14**: 161-168.
47. Nelson MS, Sadowsky MJ: Secretion systems and signal exchange between nitrogen-fixing rhizobia and legumes. Front Plant Sci. 2015, **6:** 491.
48. Meloni S, Rey L, Sidler S, Imperial J, Ruiz-Argüeso T, Palacios JM: The twin-arginine translocation (Tat) system is essential for *Rhizobium*-legume symbiosis. Molec. Microbiol. 2003, **48**: 1195-1207.
49. Deakin WJ, Broughton WJ: (2009) Symbiotic use of pathogenic strategies: rhizobial protein secretion systems. Nat. Rev. Microbiol. 2009, **7**: 312-320.
50. Alloisio N, Queiroux C, Fournier P, Pujic P, Normand P, Vallenet D, Medigue C, Yamaura M, Kakoi K, Kucho K: The *Frankia alni* symbiotic transcriptome. Mol. Plant Microbe Interact. 2010, **23**: 593-607.
51. Guefrachi I, Pierre O, Timchenko T, Alunni B, Barrière Q, Czernic P, Villaécija-Aguilar JA, Verly C, Bourge M, Fardoux J, et al.: *Bradyrhizobium* BclA is a peptide transporter required for bacterial differentiation in symbiosis with *Aeschynomene* legumes. Mol. Plant Microbe Interact. 2015, **28**: 1155-1166.
52. Kim DH, Parupalli S, Azam S, Lee SH, Varshney RK: Comparative sequence analysis of nitrogen fixation-related genes in six legumes. Front Plant Sci 2013, **4:** 300.
53. Jeong H, Mason SP, Barabasi AL, Oltvai ZN: Lethality and centrality in protein networks. Nature 2001, **411:** 41-42.
54. Han JD, Bertin N, Hao T, Goldberg DS, Berriz GF, Zhang LV, Dupuy D, Walhout AJ, Cusick ME, Roth FP, et al.: Evidence for dynamically organized modularity in the yeast protein-protein interaction network. Nature 2004, **430**: 88-93.
55. Bertolazzi P, Bock ME, Guerra C: On the functional and structural characterization of hubs in protein-protein interaction networks. Biotechnol Adv. 2013, **31**: 274-286.
56. Kiran M, Nagarajaram HA: Interaction and localization diversities of global and local hubs in human protein-protein interaction networks. Mol. Biosyst. 2016, **12**(9): 2875-2882.
57. Ungermann C, Langosch D: Functions of SNAREs in intracellular membrane fusion and lipid bilayer mixing.

J. Cell Sci. 2005, **118**: 3819-3828.

58. Hakoyama T, Oi R, Hazuma K, Suga E, Adachi Y, Kobayashi M, Akai R, Sato S, Fukai E, Tabata S, et al.: The SNARE protein SYP71 expressed in vascular tissues is involved in symbiotic nitrogen fixation in *Lotus japonicus* nodules. Plant Physiol. 2012, **160:** 897-905.

59. Catalano CM, Czymmek KJ, Gann JG, Sherrier DJ: *Medicago truncatula* syntaxin SYP132 defines the symbiosome membrane and infection droplet membrane in root nodules. *Planta*, 2007, **225**: 541-550.

60. Pan H, Oztas O, Zhang X, Wu X, Stonoha C, Wang E, Wang B, Wang D: A symbiotic SNARE protein generated by alternative termination of transcription. Nature Plants 2016, **2**: 15197.

61. Li X, Dhaubhadel S: Soybean 14-3-3 gene family: identification and molecular characterization. Planta 2011, **233**: 569-582.

62. Radwan O, Wu X, Govindarajulu M, Libault M, Neece DJ, Oh MH, Berg RH, Stacey G, Taylor CG, Huber SC, et al.: 14-3-3 proteins SGF14c and SGF14l play critical roles during soybean nodulation. Plant Physiol. 2012, **160**: 2125-2136.

63. Wienkoop S, Saalbach G: Proteome analysis. Novel proteins identified at the peribacteroid membrane from *Lotus japonicus* root nodules. Plant Physiol. 2003, **131**: 1080-1090.

64. Winzer T, Bairl A, Linder M, Linder D, Werner D, Müller P: A novel 53-kDa nodulin of the symbiosome membrane of soybean nodules, controlled by *Bradyrhizobium japonicum*. Mol. Plant Microbe Interact. 1999, **12**: 218-226.

65. Vinardell JM, Fedorova E, Cebolla A, Kevei Z, Horvath G, Kelemen Z, Tarayre S, Roudier F, Mergaert P, Kondorosi A, et al.: Endoreduplication mediated by the anaphase-promoting complex activator CCS52A is required for symbiotic cell differentiation in *Medicago truncatula* nodules. Plant Cell 2003, **15**: 2093-2105.

66. Saito K, Yoshikawa M, Yano K, Miwa H, Uchida H, Asamizu E, Sato S, Tabata S, Imaizumi-Anraku H, Umehara Y, et al.: NUCLEOPORIN85 is required for calcium spiking, fungal and bacterial symbioses, and seed production in *Lotus japonicus*. Plant Cell 2007, **19**: 610-624.

67. Kanamori N, Madsen LH, Radutoiu S, Frantescu M, Quistgaard EM, Miwa H, Downie JA, James EK, Felle HH, Haaning LL, et al.: A nucleoporin is required for induction of Ca^{2+} spiking in legume nodule development and essential for rhizobial and fungal symbiosis. Proc. Natl. Acad. Sci. U S A. 2006, **103:** 359-364.

68. Libault M, Farmer A, Brechenmacher L, Drnevich J, Langley RJ, Bilgin DD, Radwan O, Neece DJ, Clough SJ, May GD, et al.: Complete transcriptome of the soybean root hair cell, a single-cell model, and its alteration in response to *Bradyrhizobium japonicum* infection. Plant Physiol., 2010, **152**: 541-552.

69. Sugiyama A, Shitan N, Yazaki K: Involvement of a soybean ATP-binding cassette-type transporter in the secretion of genistein, a signal flavonoid in legume-*Rhizobium* symbiosis. Plant Physiol. 2007, **144**: 2000-2008.

70. Clarke VC, Loughlin PC, Day DA, Smith PM: Transport processes of the legume symbiosome membrane. Front Plant Sci. 2014, **5**: 699.

71. Colebatch G, Desbrosses G, Ott T, Krusell L, Montanari O, Kloska S, Kopka J, Udvardi MK: Global changes in transcription orchestrate metabolic differentiation during symbiotic nitrogen fixation in *Lotus*

japonicus. Plant J. 2004, **39**: 487-512.

72. Limpens E, Bisseling T: Signaling in symbiosis. Curr. Opin. Plant Biol. 2003, **6**: 343-350.
73. Choudhury SR, Pandey S: Specific subunits of heterotrimeric G proteins play important roles during nodulation in soybean. Plant Physiol., 2013, **162**: 522-533.
74. Choudhury SR, Pandey S: Phosphorylation-dependent regulation of G-protein cycle during nodule formation in soybean. Plant Cell 2015, **27**: 3260-3276.
75. Ke D, Fang Q, Chen C, Zhu H, Chen T, Chang X, Yuan S, Kang H, Ma L, Hong Z, et al.: The small GTPase ROP6 interacts with NFR5 and is involved in nodule formation in *Lotus japonicus*. Plant Physiol. 2012, **159**: 131-143.
76. Lei MJ, Wang Q, Li X, Chen A, Luo L, Xie Y, Li G, Luo D, Mysore KS, Wen J, et al.: The small GTPase ROP10 of *Medicago truncatula* is required for both tip growth of root hairs and nod factor-induced root hair deformation. Plant Cell 2015, **27:** 806-822.
77. Harper JF, Harmon A: Plants, symbiosis and parasites: a calcium signalling connection. Nat. Rev. Mol. Cell Biol. 2005, **6**: 555-566.
78. Bapaume L, Reinhardt D: How membranes shape plant symbioses: signaling and transport in nodulation and arbuscular mycorrhiza. Front Plant Sci. 2012, **3:** 223.
79. Chen T, Zhu H, Ke D, Cai K, Wang C, Gou H, Hong Z, Zhang Z: A MAP kinase kinase interacts with SymRK and regulates nodule organogenesis in *Lotus japonicus*. Plant Cell 2012, **24**: 823-838.
80. Stewart GR, Young DB: Heat-shock proteins and the host–pathogen interaction during bacterial infection. Curr. Opin. Immunol. 2004, **16**: 506-510.
81. Colditz F, Nyamsuren O, Niehaus K, Eubel H, Braun HP, Krajinski F: Proteomic approach: identification of *Medicago truncatula* proteins induced in roots after infection with the pathogenic oomycete *Aphanomyces euteiches*. Plant Mol. Biol. 2004, **55**: 109-120.
82. Oehrle NW, Sarma AD, Waters JK, Emerich DW: Proteomic analysis of soybean nodule cytosol. Phytochemistry 2008, **69:** 2426-2438.
83. Salavati A, Taleei A, Bushehri AA, Komatsu S: Analysis of the proteome of common bean (*Phaseolus vulgaris L.*) roots after inoculation with *Rhizobium etli*. Protein & Peptide Letters 2012, **19**: 880.
84. Panter S, Thomson R, de Bruxelles G, Laver D, Trevaskis B, Udvardi M: Identification with proteomics of novel proteins associated with the peribacteroid membrane of soybean root nodules. Mol. Plant Microbe Interact. 2000, **13:** 325-333.
85. Catalano CM, Lane WS, Sherrier DJ: Biochemical characterization of symbiosome membrane proteins from *Medicago truncatula* root nodules. Electrophoresis 2004, **25**: 519-531.
86. Ashby MK: Survey of the number of two-component response regulator genes in the complete and annotated genome sequences of prokaryotes. FEMS Microbiol. Lett. 2004, **231**: 277-281.
87. Whitworth DE, Cock PJ: Evolution of prokaryotic two-component systems: insights from comparative genomics. Amino Acids 2009, **37**: 459-466.
88. Charles TC, Jin S, Nester EW: Two-component sensory transduction systems in phytobacteria. Annu. Rev. Phytopathol. 1992, **30**: 463-484.
89. West AH, Stock AM: Histidine kinases and response regulator proteins in two-component signaling systems.

Trends Biochem. Sci. 2001, **26**: 369-376.

90. Hunter WJ: Increased nodulation of soybean by a strain of *Bradyrhizobium japonicum* with altered tryptophan metabolism. Lett. Appl. Microbiol. 1994, **18:** 340-342.

91. Lambrecht M, Okon Y, Vande Broek A, Vanderleyden J: Indole-3-acetic acid: a reciprocal signalling molecule in bacteria-plant interactions. Trends Microbiol. 2000, **8:** 298-300.

92. Ghosh S, Basu PS: Production and metabolism of indole acetic acid in roots and root nodules of *Phaseolus mungo*. Microbiol. Res. 2006, **161**: 362-366.

93. Roux B, Rodde N, Jardinaud MF, Timmers T, Sauviac L, Cottret L, Carrère S, Sallet E, Courcelle E, Moreau S, et al.: An integrated analysis of plant and bacterial gene expression in symbiotic root nodules using laser-capture microdissection coupled to RNA sequencing. Plant J. 2014, **77**: 817- 837.

94. Sokolovski S, Hills A, Gay RA, Blatt MR: Functional interaction of the SNARE protein NtSyp121 in Ca^{2+} channel gating, Ca^{2+} transients and ABA signalling of stomatal guard cells. Mol. Plant 2008, **1**: 347-358.

95. Lodwig E, Poole P: Metabolism of *Rhizobium* bacteroids. Crit. Rev. Plant Sci. 2003, **22**, 37-78.

96. Oke V, Long SR: Bacteroid formation in the rhizobium–legume symbiosis. Curr. Opin. Microbiol. 1999, **2**, 641-646.

97. Kersey PJ, Allen JE, Armean I, Boddu S, Bolt BJ, Carvalho-Silva D, Christensen M, Davis P, Falin LJ, Grabmueller C, et al.: Ensembl Genomes 2016: more genomes, more complexity. Nucleic Acids Res. 2016, **44:** D574-D580.

98. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, et al.: Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res. 2012, **40**: D1178-D1186.

99. Li QG, Zhang L, Li C, Dunwell JM, Zhang YM: Comparative genomics suggests that an ancestral polyploidy event leads to enhanced root nodule symbiosis in the Papilionoideae. Mol. Biol. Evol. 2013, **30**: 2602-2611.

100. Sonnhammer EL, Östlund G: InParanoid 8: orthology analysis between 273 proteomes, mostly eukaryotic. Nucleic Acids Res. 2015, **43**: D234-D239.

101. Yellaboina S, Tasneem A, Zaykin DV, Raghavachari B, Jothi R: DOMINE: a comprehensive collection of known and predicted domain-domain interactions. Nucleic Acids Res. 2011, **39**: D730-D735.

102. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, et al.: The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 2016, **44:** D279-D285.

103. O'Brien KP, Westerlund I, Sonnhammer EL: OrthoDisease: a database of human disease orthologs. Hum. Mutat. 2004, **24**, 112-119.

104. O'Brien KP, Remm M, Sonnhammer EL: Inparanoid: a comprehensive database of eukaryotic orthologs. Nucleic Acids Res. 2005, **33**: D476-D480.

105. Krogh A, Larsson B, von Heijne G, Sonnhammer EL: Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. 2001, **305:** 567-580.

106. Petersen TN, Brunak S, von Heijne G, Nielsen H: SignalP 4.0: discriminating signal peptides from

954 transmembrane regions. Nat. Methods 2011, **8**: 785-786.
955 107. Huntley RP, Sawford T, Mutowo-Meullenet P, Shypitsyna A, Bonilla C, Martin MJ, O'Donovan C: The
956 GOA database: gene Ontology annotation updates for 2015. Nucleic Acids Res. 2015, **43**, D1057- D1063.
957 108. Jiang JJ, Conrath DW: *Semantic similarity based on corpus statistics and lexical taxonomy*. In Taiwan:
958 Proceedings of International Conference Research on Computational Linguistics (ROCLING X), 1997.
959 109. Schlicker A, Albrecht M: FunSimMat: a comprehensive functional similarity database. Nucleic Acids Res.
960 2008, **36**: D434-D439.
961 110. Schlicker A, Domingues FS, Rahnenführer J, Lengauer T: A new measure for functional similarity of gene
962 products based on Gene Ontology. BMC Bioinformatics 2006, **7**: 302.
963 111. Couto FM, Silva MJ, Coutinho PM: Measuring semantic similarity between Gene Ontology terms. Data &
964 Knowledge Engineering 2007, **61**: 137-152.
965 112. Casse F, Boucher C, Julliot JS, Michel M, Dénarié J: Identification and characterization of large plasmids in
966 *Rhizobium meliloti* using Agarose Gel Electrophoresis. Journal of General Microbiology 1979 **113**: 229-242.
967 113. Krenek P, Samajova O, Luptovciak I, Doskocilova A, Komis G, Samaj J: Transient plant transformation
968 mediated by Agrobacterium tumefaciens: Principles, methods and applications. Biotechnol. Adv. 2015, **33:**
969 1024-1042.
970 114. Walter M, Chaban C, Schütze K, Batistic O, Weckermann K, Näke C, Blazevic D, Grefen C, Schumacher K,
971 Oecking C, Harter K, Kudla J: Visualization of protein interactions in living plant cells using bimolecular
972 fluorescence complementation. Plant J. 2004, **40**: 428-438.
973 115. Wang J, Cheng G, Wang C, He Z, Lan X, Zhang S, Lan H: The bHLH transcription factor CgbHLH001 is a
974 potential interaction partner of CDPK in halophyte *Chenopodium glaucum*. Sci. Rep. 2017, **7**(1): 8441

975 **Table 1 Classification of proteins in predicted PPIs between soybean and *B. diazoefficiens* USDA 110 by PANTHER overrepresentation test**

PANTHER Protein Class	Observed	Expected	Fold Enrichment	Corrected P-value	PANTHER Protein Class	Observed	Expected	Fold Enrichment	Corrected P-value
Bradyrhizobium diazoefficiens USDA 110					membrane traffic protein	84	28.82	2.92	4.94E-15
carbohydrate transporter	6	0.32	> 5	1.18E-04	vesicle coat protein	16	5.53	2.89	3.56E-02
cation transporter	26	1.61	> 5	5.00E-21	amino acid transporter	35	14.62	2.39	6.70E-04
ion channel	5	0.47	> 5	1.26E-02	transfer/carrier protein	74	31.73	2.33	1.28E-08
transporter	68	15.96	4.26	2.23E-22	transporter	221	121.95	1.81	9.10E-15
Glycine max					**III. metabolism**				
I. gene transcription and translation					ATP synthase	18	2.83	> 5	2.39E-07
deacetylase	13	2.23	> 5	1.15E-04	oxidase	29	11.79	2.46	2.66E-03
aminoacyl-tRNA synthetase	12	2.70	4.44	4.47E-03	reductase	64	28.22	2.27	6.73E-07
ribosomal protein	97	30.62	3.17	6.74E-20	enzyme modulator	120	57.55	2.09	3.26E-11
translation initiation factor	25	7.98	3.13	1.79E-04	dehydrogenase	79	38.64	2.04	9.97E-07
RNA helicase	22	7.33	3	1.50E-03	isomerase	49	24.06	2.04	7.95E-04
translation factor	44	15.35	2.87	2.71E-07	oxidoreductase	176	104.54	1.68	6.28E-09
translation elongation factor	20	7.46	2.68	1.71E-02	hydrolase	221	136.49	1.62	6.11E-10
helicase	28	11.23	2.49	2.93E-03	ligase	69	42.79	1.61	2.04E-02
RNA binding protein	234	127.4	1.84	2.15E-16	transferase	236	181.6	1.3	5.17E-03
chaperone	54	30.19	1.79	8.93E-03	**IV. signaling**				
nucleic acid binding	355	205.87	1.72	2.06E-21	G-protein	34	10.51	3.24	1.07E-06
II. transport and intracellular trafficking					small GTPase	18	6.30	2.86	1.74E-02
anion channel	10	1.54	> 5	8.85E-04	G-protein modulator	36	14.84	2.43	3.66E-04
ATP-binding cassette (ABC) transporter	41	8.02	> 5	1.92E-14	calcium-binding protein	49	20.84	2.35	1.51E-05
mitochondrial carrier protein	26	6.26	4.15	4.99E-07	intracellular calcium-sensing protein	32	14.97	2.14	1.39E-02
cation transporter	49	13.85	3.54	2.45E-11	calmodulin	32	14.97	2.14	1.39E-02
membrane trafficking regulatory protein	18	5.32	3.39	2.02E-03	SNARE protein	18	5.02	3.59	9.36E-04
ion channel	26	8.53	3.05	1.84E-04					

976

977

978 **Table 2 KEGG pathway enrichment analysis of proteins in PPIs between soybean and *B. diazoefficiens* USDA 110**

Glycine max					***Bradyrhizobium diazoefficiens* USDA 110**				
KEGG Term	**KEGG ID**	**Input number**	**Background number**	**Corrected P-Value**	**KEGG Term**	**KEGG ID**	**Input number**	**Background number**	**Corrected P-Value**
Oxidative phosphorylation	gmx00190	72	237	9.35E-09	Oxidative phosphorylation	bja00190	20	66	1.46E-10
Phagosome	gmx04145	52	165	7.00E-07	Protein export	bja03060	10	20	5.58E-07
Protein export	gmx03060	36	91	8.57E-07	Two-component system	bja02020	20	168	5.79E-05
Protein processing in endoplasmic reticulum	gmx04141	81	375	6.27E-05	Peptidoglycan biosynthesis	bja00550	6	24	0.003838
N-Glycan biosynthesis	gmx00510	27	74	9.75E-05	Glycerophospholipid metabolism	bja00564	6	24	0.003838
Ribosome	gmx03010	109	595	0.000594	ABC transporters	bja02010	22	299	0.007615
Biosynthesis of amino acids	gmx01230	77	429	0.011763	Bacterial secretion system	bja03070	7	44	0.009047
Citrate cycle (TCA cycle)	gmx00020	26	104	0.015133	beta-Lactam resistance	bja01501	5	21	0.009047
Carbon metabolism	gmx01200	84	488	0.016097					
Proteasome	gmx03050	24	105	0.049615					

979 Notes: Input genes and their detailed annotations were available in Supplementary Table S6

980

Table 3 Top ten hubs of *G. max* and *B. diazoefficiens* USDA 110 in the predicted PPI network

Glycine max			***Bradyrhizobium diazoefficiens* USDA 110**		
Gene	**Gene annotation**	**Degree**	**Gene**	**Gene annotation**	**Degree**
Glyma.14G176900	14-3-3 protein (SGF14k)	33	*BAC49080*	putative cation-transporting ATPase (EC 3.6.3.-)	347
Glyma.02G208700	14-3-3 protein (SGF14g)	33	*BAC52411*	metalloprotease	300
Glyma.04G102900	pumilio 7	21	*BAC49957*	peptidyl prolyl cis-trans isomerase	172
Glyma.08G332900	heat shock protein 81.4	17	*BAC52381*	aquaporin Z	155
Glyma.13G359500	heat shock protein 91	17	*BAC45806*	hypothetical protein	137
Glyma.18G074100	heat shock protein 81.4	17	*BAC45833*	glycerol-3-phosphate dehydrogenase [NAD(P)$^+$]	124
Glyma.15G014400	heat shock protein 91	17	*BAC47677*	hypothetical protein	109
Glyma.12G116300	ADP/ATP carrier 3	16	*BAC47750*	rieske iron-sulfur protein	106
Glyma.06G290600	ADP/ATP carrier 3	16	*BAC45992*	hypothetical protein	102
Glyma.10G193200	heat shock protein 60	16	*BAC46205*	putative Na^+/H^+ exchanger	95

983 **Figure legends**

984

985 **Figure 1. The prediction pipeline of the protein-protein interaction networks**

986

987 **Figure 2. Distribution of semantic similarity scores between GO terms of two**
988 **proteins: sim_{JC}^{BP}, sim_{JC}^{MF} and sim_{JC}^{CC}**. A: distribution of sim_{JC}^{BP}; B: distribution of
989 sim_{JC}^{CC}; C: distribution of sim_{JC}^{MF}. Box in black represents predicted protein-protein
990 interactions in soybean; grey box denotes random protein pairs in the soybean
991 genome.

992

993 **Figure 3. Two PPI sub-networks between soybean (red) and *B. diazoefficiens***
994 **USDA 110 (yellow) proteins**. A: PPI sub-network between 18 soybean SNARE
995 proteins and *B. diazoefficiens* USDA 110 proteins. B: PPI sub-network of DctA and
996 14-3-3 proteins. Triangles represent nodulin in soybean. The PPI interactions with
997 bold edges were validated by the LCI assay.

998

999 **Figure 4. Luciferase complementation image assay of a subnetwork containing**
1000 **two 14-3-3 proteins in *Agrobacterium*-infiltrated *N. benthamiana* leaves under**
1001 **bright field (I) and dark (II) illumination**. The C-terminal half and the N-terminal
1002 half of LUC were fused to N-gene, N-LUC, C-gene and C-LUC. In (c), the treatment
1003 was N-GmSGF14g + C-BAC48988, and the controls were N-LUC + C-BAC48988,
1004 N-GmSGF14g + C-LUC, and N-LUC + C-LUC. LUC fluorescence was detected by
1005 confocal microscope in *N. benthamiana* fresh leaves. The experiment was repeated
1006 three times with similar results. The situation was similar in the others.