

Bradyrhizobium diazoefficiens USDA 110–
glycine max interactome provides
candidate proteins associated with
symbiosis

Article

Accepted Version

Zhang, L., Liu, J.-Y., Gu, H., Du, Y., Zuo, J.-F., Zhang, Z., Zhang, M., Li, P., Dunwell, J. M. ORCID: <https://orcid.org/0000-0003-2147-665X>, Cao, Y., Zhang, Z. and Zhang, Y.-M. (2018) *Bradyrhizobium diazoefficiens* USDA 110–glycine max interactome provides candidate proteins associated with symbiosis. *Journal of Proteome Research*, 17 (9). pp. 3061-3074. ISSN 1535-3893 doi: <https://doi.org/10.1021/acs.jproteome.8b00209> Available at <https://centaur.reading.ac.uk/78921/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1021/acs.jproteome.8b00209>

Publisher: American Chemical Society

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in

the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

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

4
5 Li Zhang^{1,†}, Jin-Yang Liu^{2,†}, Huan Gu², Yanfang Du³, Jian-Fang Zuo³,
6 Zhibin Zhang³, Menglin Zhang³, Pan Li⁵, Jim M. Dunwell⁶,
7 Yangrong Cao⁷, Zuxin Zhang^{4,*} and Yuan-Ming Zhang^{1,*}

8
9 1 College of Plant Science and Technology, Huazhong Agricultural University, Wuhan 430070,
10 China / Xinxiang Key Laboratory of Public Health Informatics, School of Public Health,
11 Xinxiang Medical University, Xinxiang 453003, China

12 2 College of Agriculture, Nanjing Agricultural University, Nanjing 210095, China

13 3 College of Plant Science and Technology, Huazhong Agricultural University, Wuhan 430070,
14 China

15 4 National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University,
16 Wuhan 430070, China

17 5 Xinxiang Key Laboratory of Public Health Informatics, School of Public Health, Xinxiang
18 Medical University, Xinxiang 453003, China

19 6 School of Agriculture, Policy and Development, University of Reading, Reading RG6 6AR,
20 United Kingdom

21 7 College of Life Science and Technology, Huazhong Agricultural University, Wuhan 430070,
22 China

23
24 †: These authors contributed equally to this work.

25
26 * **Correspondences** (e-mails soy Zhang@mail.hzau.edu.cn or zuxin Zhang@mail.hzau.edu.cn)

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

28
29 **Data Availability Statement** All the datasets analyzed were from previously
30 published datasets. Supporting Information may be found in additional files.

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

36

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

39

40 **Abstract**

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

61

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

64

65 **Introduction**

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

72
73 Among rhizobia, *Bradyrhizobium diazoefficiens* USDA 110 (previously named
74 *Bradyrhizobium japonicum* USDA 110) is the most agriculturally important rhizobial
75 bacterium as it is able to specifically infect soybean (*Glycine max*), one of the
76 important legume plants in the world, and form a nitrogen-fixing symbiosis [3].
77 Furthermore, *G. max*-*B. diazoefficiens* is one of the most studied soybean-rhizobium
78 symbiotic models [4]. Given the importance of such unique feature of legumes,
79 further studies on the mechanisms of the soybean-rhizobium symbiosis are of
80 particular interest. Importantly, the genome sequences of both *B. diazoefficiens* USDA
81 110 and *G. max* are now available [3,5], and provide an opportunity to better
82 understand the mechanism of symbiotic features in terms of genomics and
83 proteomics.

84
85 In *B. diazoefficiens* USDA 110, several genes related to various stages of the
86 symbiosis process have been identified [3]. In soybean, comparative genomics
87 analysis of legumes also predicted several nodulin genes [5]. Additionally, microarray
88 approaches and RNA-seq analysis in soybean revealed a large number of genes
89 differentially regulated during the symbiosis [4,6,7]. However, none of the above
90 studies have focused on the complex interactions between candidate symbiosis-related
91 genes. Generally, the proteins in the symbiosis process function as a complex network,
92 which combines complex chemical, physical and biological interactions between
93 rhizobial bacteria and their host plants [8]. To better elucidate the complex microbial

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

97

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

112

113 In the past decade, a series of computational approaches for PPI prediction have been
114 developed [16,17], and these now play important roles in complementing the various
115 experimental approaches. The existing computational approaches for PPI prediction
116 have exploited diverse data features, which include domain and motif information
117 [18-21], network topology [21,22], gene ontology (GO) [18-20], gene expression
118 [18,19], protein sequence similarity [14,23], and pathway analysis [24]. At present,
119 the interolog and domain-based approaches [25-27] are widely used [14,15,28]. The
120 interolog method is based on protein sequence similarity to conduct the PPI prediction,
121 which maps interactions in the source organism onto the target organism to find
122 possible interactions in the target organism [25,26]. The domain-based method uses

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

126

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

138

139 **Results**

140 **Network construction**

141

142 Based on the well-studied experimental PPIs of seven model organisms: *Arabidopsis*
143 *thaliana*, *Caenorhabditis elegans*, *Drosophila melanogaster*, *Escherichia coli* K12,
144 *Homo sapiens*, *Mus musculus* and *Saccharomyces cerevisiae*, the PPIs between *G.*
145 *max* and *B. diazoefficiens* were predicted in this study. To make use of more
146 comprehensive information, we obtained the PPIs of seven organisms from multiple
147 databases: BioGrid [31], DIP [32], HPRD [33], IntAct [34], MINT [31] and TAIR
148 [35]. An ID dictionary was obtained from BioGrid to provide cross-database ID
149 mapping. For mismatching IDs, we corrected manually in the Uniprot ID mapping
150 server. As a result, we incorporated 44702 PPIs with 9948 proteins in *A. thaliana*,
151 28791 PPIs with 11543 proteins in *C. elegans*, 78383 PPIs with 9438 proteins in *D.*

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

155

156 All the 56044 *G. max* and 8317 *B. diazoefficiens* USDA 110 proteins were used to
157 conduct a genome-wide PPI prediction. Among 8317 *B. diazoefficiens* proteins, 2356
158 proteins are secreted or membrane proteins (Table S2), which have the possibility to
159 interact with *G. max* proteins. Using the pipeline shown in Figure 1 and filtered by
160 above 2356 secreted or membrane proteins, 5115 PPIs between 2291 soybean proteins
161 and 290 *B. diazoefficiens* USDA 110 proteins were predicted (Figure S1; Table S3). In
162 addition, 233545 intra-species PPIs in soybean (Table S4) and 11106 intra-species
163 PPIs in *B. diazoefficiens* USDA 110 (Table S5) were predicted. In summary, there
164 were a total of 249766 PPIs, including inter- and intra-species PPIs, and 54471 PPIs
165 (21.81%) were found in more than one species or experiment. All predicted
166 interactions and the detailed annotation information of the proteins are available in
167 Tables S3 to S5.

168

169 **Quality assessment of protein–protein interactions**

170

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

180

181 **Expression pattern in nodules** The interaction between rhizobium and its host

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

193

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

199

200 **Functional similarity based on GO annotation** Two interacting proteins would
201 have similar or related functions and should share some common GO annotations
202 [23,28,36]. Thus, GO annotation information of two interacting proteins was used to
203 measure the accuracy of our prediction. Among 56044 soybean genes, 30023 (53.57%)
204 genes were annotated with at least one GO term in any of the three GO categories
205 (molecular function, biological process, and cellular component). Of all the 233545
206 soybean PPIs, 128862, 66369 and 26135 PPIs were annotated in the categories of
207 molecular function, biological process, and cellular component (125086, 63581,
208 25007 non-self interactions), respectively (Table S4).

209

210 To measure the semantic similarity between GO terms and to evaluate the reliability
211 of predicted PPIs, three functional similarity scores, $\text{sim}_{\text{JC}}^{\text{BP}}$, $\text{sim}_{\text{JC}}^{\text{MF}}$ and $\text{sim}_{\text{JC}}^{\text{CC}}$,

212 were calculated using non-self interactions in each GO category. Meanwhile,
213 randomly selected protein pairs of the same size served as a control. As a result,
214 significant differences for each of three sim_{JC} scores between predicted PPIs and
215 randomly selected protein pairs were observed (Figure 2). All the proportions of score
216 1.0 in $\text{sim}_{\text{JC}}^{\text{BP}}$, $\text{sim}_{\text{JC}}^{\text{MF}}$, and $\text{sim}_{\text{JC}}^{\text{CC}}$ were significantly higher in predicted soybean
217 PPIs than those in randomly selected protein pairs, indicating that the predicted
218 interaction network indeed preferentially connects functionally related proteins.

219

220 **Co-expression of predicted soybean PPIs** Levels of mRNA expression have
221 some relationship with protein-protein interactions [42]. The interacting proteins tend
222 to have correlated gene expression patterns, especially for subunits of the same
223 protein complex [28,37,43]. Thus, we investigated the relationship of our predicted
224 intra-species PPIs with mRNA expression levels in soybean. In this study, we used the
225 transcriptome data from nine tissues of *G. max* to investigate expression correlation
226 between two interacting proteins. The co-expression level of two interacting proteins
227 was calculated by a widely used measure, the Pearson correlation coefficient (PCC)
228 [44].

229

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

239

240 **Conserved PPIs identified in more than two species**

241

242 Common protein interactions predicted from multiple species can be considered as
243 evolutionarily conserved interactions that have very high confidence [36]. In this
244 study, we detected common protein interactions from more than two species. As a
245 result, 60 conserved PPIs including 54 *G. max* proteins and 21 *B. diazoefficiens*
246 proteins in *G. max*-*B. diazoefficiens* interactome were found (Figure S2). Among these
247 54 *G. max* proteins, more importantly, 49 proteins were expressed with FPKM > 5 in
248 the underground tissues (root, root hair and nodule) and 24 proteins had high
249 expression levels with FPKM > 100 in the underground tissues.

250

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

252

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

258

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

269

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

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

290

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

292

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

301 hubs, we performed KEGG pathway enrichment analysis for the proteins interacting
302 with each of the twenty hubs. These results are listed in Supplementary [Table S10](#).

303

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

313

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

319

320 **Subnetworks related to symbiosis**

321

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

327

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

332 involved in the predicted *G. max*-*B. diazoefficiens* interactome and closely interacted
333 with *B. diazoefficiens* proteins (Figure 3A), suggesting the critical roles of SNAREs in
334 soybean root nodule symbiosis (RNS). Meanwhile, soybean SNAREs interacted with
335 each other in Figure 3A, which was consistent with the results in previous structural
336 studies that SNAREs could form complexes by interacting with other SNAREs
337 [57,58].

338
339 **14-3-3 protein** 14-3-3 proteins are abundant proteins in plants, and are involved
340 in signaling pathways to regulate plant development and response to stimulus. Li and
341 Dhaubhadel [61] identified 18 genes (SGF 14a-r) coding 14-3-3 proteins in the whole
342 soybean genome. Previous studies revealed that two of them (SGF14c and SGF14l)
343 play critical roles in RNS [62] and homologs of SGF14b in *L. japonicus* were located
344 in the peribacteroid membrane [63]. In our study, we found another two 14-3-3
345 proteins, Glyma.14G176900 (SGF14k) and Glyma.02G208700 (SGF14g), which are
346 hubs that interacted with *B. diazoefficiens* to a high degree (Table 3). More
347 importantly, we found that SGF14k and SGF14g were connected with four soybean
348 nodulation genes, *Glyma.06G065600* (Nodulin26) [64], *Glyma.17G13300* (WD40
349 protein; homologs of MtCCS52) [65], *Glyma.17G193800* (nucleoporin; homologs of
350 LjNUP85) [66] and *Glyma.14G008200* (nucleoporin; homologs of LjNUP133)
351 (Figure 3B) [67]. The results of the predicted PPIs of SGF14k and SGF14g
352 demonstrated that SGF14k and SGF14g were involved in RNS.

353
354 **Validation of a subnetwork containing two 14-3-3 proteins using luciferase**
355 **complementation image (LCI) assay experiment**

356
357 Luciferase complementation image (LCI) assay is a well-established method to verify
358 the predicted PPIs in a laboratory setting. To validate the accuracy of the predicted
359 interactions, a subnetwork in Figure 3B was selected to test the interactions *in vivo*.
360 As a result, nine were confirmed (Figure 4). For example, SGF14k was interacted
361 with BAC48988, similarly, SGF14k and BAC49563, SGF14k and Nodulin26,

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

374

375 **Discussion**

376 **Network validation for the predicted PPIs in this study**

377

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

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

393

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

396

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

419 involvement of many nod factors in the signal transduction processes such as
420 G-protein coupled receptor signaling pathways [73,74], small GTPase mediated signal
421 transduction [75,76], calmodulin [77], Soluble N-Ethylmaleimide Sensitive Factor
422 Attachment Protein Receptor (SNARE) proteins [58,78] and the MAPK
423 (Mitogen-activated protein kinase) cascade [79]. In the present study, 34 G-proteins,
424 18 small GTPase, 32 calmodulin and 18 SNARE proteins were present in the
425 predicted PPIs and directly interacted with bacteria (Tables 1 and S8). The
426 subnetworks of related signaling transduction provide opportunities to reveal whether
427 and how these networks are interconnected, and then give insights into the mechanism
428 of symbiosis.

429

430 **Hubs in the predicted network played roles in symbiosis**

431

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

444

445 The other two highly interacting hubs were SGF14k and SGF14g, which were shown
446 to interact with *B. diazoefficiens* proteins in the pathways of two-component systems
447 (TCSs) and tryptophan metabolism (Table S10). TCSs are abundant signaling
448 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

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

480

481 **Subnetwork analysis provide insight into the mechanism of root nodule**
482 **symbiosis**

483

484 Subnetwork analysis of SNAREs showed that SNAREs mainly interacted with
485 membrane transporters or related proteins (Figure 3A). In detail, Glyma.10G008300
486 interacted with a cation efflux system protein (BAC50315), two ABC transporter
487 permease proteins (BAC51159 and BAC49765), a cation-transporting ATPase
488 (BAC52318) and an ammonium transporter (BAC45878). Five SNARE proteins
489 (Glyma.04G072700, Glyma.10G149000, Glyma.07G042400, Glyma.03G029700 and
490 Glyma.01G137300) interacted with BAC49080, a cation-transporting ATPase.
491 Glyma.10G149000 and Glyma.13G307600 interacted with a Na⁺/H⁺ exchanger
492 (BAC46205). Sokolovski *et al.* [94] proved that a plasma membrane SNARE protein
493 in *Nicotiana benthamiana* guard cells could regulate Ca²⁺ channels and also possibly
494 target other ion channels. The results indicated that SNAREs in the symbiosome
495 membrane may play roles in regulating bacteria ion channels. Further analysis of the
496 role of SNARE proteins will provide novel insights into RNS.

497

498 Through the subnetwork analysis, two 14-3-3 proteins, SGF14k and SGF14g, not only
499 interacted with soybean nodulins but also were closely connected with two bacterial
500 DctA proteins, BAC48988 and BAC49563 (Figure 3B). DctA was an important
501 transporter for C4-dicarboxylic acids, which are the main form of carbon and energy
502 sources from host plant to *rhizobium* [95]. Notably, DctA was reported to be essential
503 for symbiotic nitrogen fixation in *Sinorhizobium meliloti*, as well as other rhizobia
504 [78,96]. The relationships between the above two 14-3-3 proteins and DctA proteins
505 were further verified by LCI assay experiments (Figure 4). Taken together, the results
506 indicated that 14-3-3 proteins SGF14g and SGF14k regulate rhizobium DctA.

507

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

515

516 **Materials and methods**

517 **Datasets**

518

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

527

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

536

537 To carry out the domain-based PPI prediction, we downloaded the interacting Pfam
538 domain pairs from the database of protein domain interactions (DOMINE Version 2.0)
539 [101], which contains a total of 26219 domain-domain interactions (DDI). To increase
540 the accuracy of prediction, only 2989 high-confident domain pairs were used as
541 reference in this study.

542

543 **PPI prediction**

544

545 Our PPI prediction was mainly based on the interolog method, along with the
546 domain-based method to improve prediction accuracy. In the interolog method
547 proposed by Walhout *et al.* [25], the pair of interactions A–B and A1–B1 are called an
548 interolog if interacting proteins A and B in a species have interacting orthologs A1 and
549 B1 in another species. Based on this theory, interolog PPI prediction is a process that
550 maps interactions in the source organism onto the target organism to find possible
551 interactions [26]. In the domain-based method, the two proteins are expected to
552 interact with each other if a protein pair contains at least one interacting domain pair
553 [27]. The protein domain annotations for *B. diazoefficiens* USDA 110 were conducted
554 in the Pfam website [102], and the annotations for *G. max* proteins were obtained from
555 Phytozome V10.3.

556

557 In this study, ortholog pairs between each of the aforementioned seven model
558 organisms and *G. max* (or *B. diazoefficiens*) were obtained from the InParanoid
559 database [100]. InParanoid scores between 0 and 1 reflect the relative evolutionary
560 distance between orthologous gene pairs [103,104]. The top score 1.0 is the best blast
561 hit and has high credibility, and orthologs with scores below 1.0 are more or less
562 sensitive. To restrict the sensitivity, ortholog pairs were selected with a score cutoff of
563 0.5. These ortholog pairs were further divided into two groups according to
564 InParanoid score: orthologs with top score 1.0 and ones with a score between 0.5 and
565 1.0. Using the interolog method, all the above ortholog pairs were mapped onto the
566 integrated PPI interactomes of the seven model organisms to predict PPIs. The

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

570

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

572

573 The transmembrane and secreted proteins in *B. diazoefficiens* USDA 110 are
574 considered to be positive candidates for interactions with *G. max*. All the proteins in *B.*
575 *diazoefficiens* USDA 110 were used to predict transmembrane proteins through
576 TMHMM 2.0 [105] and to identify secretory proteins through SingleIP 4.0 [106]. In
577 TMHMM 2.0, the proteins were inferred to be transmembrane if the number of
578 predicted transmembrane helices was not <1, and the expected number of amino acids
579 in at least one transmembrane helix was not <18. SingleIP 4.0 was employed with the
580 default settings.

581

582 **GO annotation and measurement of functional similarity**

583

584 The GO annotations of *B. diazoefficiens* USDA 110 and *G. max* were obtained from
585 the Gene Ontology Annotation (UniProt-GOA) Database [107] and Phytozome V10.3
586 [98], respectively. Semantic similarity scores between GO terms were measured by
587 Jiang and Conrath's distance method and calculated in database FunSimMat [108,109]
588 to evaluate the reliability of the predicted PPIs [110]. Jiang and Conrath's distance
589 between two GO terms is based on information content and was defined as follows
590 [111]:

$$591 \quad sim_{JC}(t_1, t_2) = \frac{1}{IC(t_1) + IC(t_2) - 2 \times IC(MIA) + 1}$$

592 $sim_{JC}(t_1, t_2)$ is the set of common ancestors of terms t_1 and t_2 in the ontology, and
593 ranges between 0, for no similarity, to 1, for highest similarity. We used sim_{JC} for
594 referring to this score. As GO annotation classifies functions of a protein according to
595 three features: molecular function, biological process and cellular component, there

596 were, correspondingly, three independent sim_{JC} scores: $\text{sim}_{\text{JC}}^{\text{MF}}$, $\text{sim}_{\text{JC}}^{\text{BP}}$ and $\text{sim}_{\text{JC}}^{\text{CC}}$.
597

598 **Co-expression analysis**

599

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

607

608 **Luciferase Complementation Image (LCI) assays for PPIs in *Nicotiana*** 609 ***benthamiana* cells**

610

611 **Materials** Soybean (*G. max* Willimas 82) and tobacco plants were grown at
612 16-hlight / 8-h dark at 25 °C for 30-60 d. *B. japonicum* (USDA110) was grown on
613 (HM) medium plates at containing 50 µg of chloramphenicol/ml for selection of
614 plasmid 25 °C.

615

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

622

623 **Primers and conditions for PCR** Primers were analyzed by Oligo 6 ([Table S11](#)).
624 PCR was carried out using a PCR system for 35 cycles (30 s at 95 °C, 30 s at Tm and
625 1-4 mins at 72 °C).

626

627 **Luciferase Complementation Image (LCI) assays** Full length coding sequence

628 of target genes were amplified by polymerase chain reaction from total RNA (Table
629 S11) and were cloned into the *Bam*HI and *Sal*I sites of JW-771-N (NLUC), as well as
630 *Kpn*I and *Sal*I sites of JW-772-C, to produce target gene-NLUC and target
631 gene-CLUC recombination vectors for the LCI assay (for split Luc
632 N-terminal/C-terminal fragment expression), respectively. Thus, N-gene, C-gene,
633 N-LUC, and C-LUC were constructed according to previously described protocols.
634 These constructs were transformed into *Agrobacterium tumefaciens* GV3101 strain
635 through CaCl₂ transformation [113]. The p19 protein (tomato bushy stunt virus) was
636 used to suppress gene silencing [114].

637

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

653

654 **Supporting information**

655

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

657 **databases**

658

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

660

661 **S3 Table. The predicted *G. max*-*B. diazoefficiens* interactome and detailed annotation**

662 **information of the proteins, including 5115 inter-species PPIs between 2291 *G. max* and 290**

663 ***B. diazoefficiens* USDA 110 proteins**

664

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

666

667 **S5 Table. The predicted *B. diazoefficiens* USDA 110 interactome, including 11106**

668 **intra-species PPIs in *B. diazoefficiens* USDA 110**

669

670 **S6 Table. List of 172 genes in the predicted PPIs that were detected to be expressed in**

671 **bacteroids of the root nodule during symbiosis in at least one of three previous studies**

672

673 **S7 Table. List of input genes enriched in KEGG pathway enrichment analysis in Table 2 and**

674 **their detailed annotation**

675

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

677

678 **S9 Table. Nodulation-related genes that experimentally interacted with *B. diazoefficiens***

679 **USDA 110 proteins**

680

681 **S10 Table. Top ten hubs of *G. max* and *B. diazoefficiens* USDA 110 in the *G. max*-*B.***

682 ***diazoefficiens* interactome and KEGG pathway enrichment analysis of these hubs by using**

683 **their interacted proteins in the PPI interactome**

684

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

686 ***Nicotiana benthamiana* cells**

687

688 **S1 Figure. Visualization of the predicted PPI network between soybean and *B. diazoefficiens***

689 **USDA 110. Each node represents a protein and each edge denotes an interaction. Red color circles**

690 **represent soybean and yellow represent *B. diazoefficiens* USDA 110.**

691

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.

- 725 9. Afroz A, Zahur M, Zeeshan N, Komatsu S: Plant-bacterium interactions analyzed by proteomics. *Front Plant*
726 *Sci.* 2013, **4**: 21.
- 727 10. Qi Y, Noble WS: *Protein interaction networks: protein domain interaction and protein function prediction*.
728 Springer Berlin Heidelberg, 2011.
- 729 11. Dyer MD, Neff C, Dufford M, Rivera CG, Shattuck D, Bassaganya-Riera J, Murali TM, Sobral BW: The
730 human-bacterial pathogen protein interaction networks of *Bacillus anthracis*, *Francisella tularensis*, and
731 *Yersinia pestis*. *PLoS ONE* 2010, **5**: e12089
- 732 12. Durmus Tekir SD, Ülgen KÖ: Systems biology of pathogen-host interaction: networks of protein-protein
733 interaction within pathogens and pathogen-human interactions in the post-genomic era. *Biotechnol. J.* 2013,
734 **8**: 85-96.
- 735 13. Martinez F, Rodrigo G, Aragonés V, Ruiz M, Lodewijk I, Fernandez U, Elena SF, Daros JA: Interaction
736 network of tobacco etch potyvirus NIa protein with the host proteome during infection. *BMC Genomics*
737 2016, **17**: 87.
- 738 14. Li ZG, He F, Zhang Z, Peng YL: Prediction of protein-protein interactions between *Ralstonia solanacearum*
739 and *Arabidopsis thaliana*. *Amino Acids* 2012, **42**: 2363-2371.
- 740 15. Sahu SS, Weirick T, Kaundal R: Predicting genome-scale *Arabidopsis-Pseudomonas syringae* interactome
741 using domain and interolog-based approaches. *BMC Bioinformatics* 2014, **15**: S13.
- 742 16. Nourani E, Khunjush F, Durmus S: Computational approaches for prediction of pathogen-host
743 protein-protein interactions. *Front Microbiol.* 2015, **6**: 94.
- 744 17. Shen J, Zhang J, Luo X, Zhu W, Yu K, Chen K, Li Y, Jiang H: Predicting protein-protein interactions based
745 only on sequences information. *Proc. Natl. Acad. Sci. U S A.* 2007, **104**: 4337- 4341.
- 746 18. Kshirsagar M, Carbonell J, Klein-Seetharaman J: Techniques to cope with missing data in host-pathogen
747 protein interaction prediction. *Bioinformatics* 2012, **28**: i466-i472.
- 748 19. Kshirsagar M, Carbonell J, Klein-Seetharaman J: Multitask learning for host-pathogen protein interactions.
749 *Bioinformatics* 2013, **29**: 217-226.
- 750 20. Coelho ED, Arrais JP, Matos S, Pereira C, Rosa N, Correia MJ, Barros M, Oliveira JL: Computational
751 prediction of the human-microbial oral interactome. *BMC Systems Biol.* 2014, **8**: 1-12.
- 752 21. Tastan O, Yanjun QI, Carbonell JG, Kleinseetharaman J: Prediction of interactions between HIV-1 and
753 human proteins by information integration. *Pacific Symposium on Biocomputing Pacific Symposium on*
754 *Biocomputing* 2015, **527**: 516-527.
- 755 22. Qi Y, Tastan O, Carbonell JG, Klein-Seetharaman J, Weston J: Semi-supervised multi-task learning for
756 predicting interactions between HIV-1 and human proteins. *Bioinformatics* 2011, **26**: i645- i652.
- 757 23. Gu H, Zhu P, Jiao Y, Meng Y, Chen M: PRIN: a predicted rice interactome network. *BMC Bioinformatics*
758 2011, **12**: 161.
- 759 24. Wuchty S: Computational prediction of host-parasite protein interactions between *P. falciparum* and *H.*
760 *sapiens*. *PLoS ONE* 2011, **6**: e26960.
- 761 25. Walhout AJM, Sordella R, Lu XW, Hartley JL, Temple GF, Brasch MA, Thierry-Mieg N, Vidal M: Protein
762 interaction mapping in *C. elegans* using proteins involved in vulval development. *Science* 2000, **287**:

- 763 116-122.
- 764 26. Matthews LR, Vaglio P, Reboul J, Ge H, Davis BP, Garrels J, Vincent S, Vidal M: Identification of potential
765 interaction networks using sequence-based searches for conserved protein-protein interactions or "interologs".
766 Genome Res. 2001, **11**: 2120-2126.
- 767 27. Flórez AF, Park D, Bhak J, Kim BC, Kuchinsky A, Morris JH, Espinosa J, Muskus C: Protein network
768 prediction and topological analysis in *Leishmania major* as a tool for drug target selection. BMC
769 Bioinformatics 2010, **11**: 484.
- 770 28. Lei D, Lin R, Yin C, Li P, Zheng A: Global protein-protein interaction network of rice sheath blight
771 pathogen. J. Proteome Res. 2014, **13**: 3277-3293.
- 772 29. Kerppola TK: Complementary methods for studies of protein interactions in living cells. Nature Methods
773 2006, **3**(12), 969-971.
- 774 30. Kerppola TK: Visualization of molecular interactions by fluorescence complementation. Nat Rev Mol. Cell
775 Biol. 2006, **7**(6): 449-456.
- 776 31. Chatr-Aryamontri A, Breitkreutz BJ, Oughtred R, Boucher L, Heinicke S, Chen D, Stark C, Breitkreutz A,
777 Kolas N, O'Donnell L, et al.: The BioGRID interaction database: 2015 update. Nucleic Acids Res. 2015, **43**,
778 D470-D478.
- 779 32. Xenarios I, Salwinski L, Duan XJ, Higney P, Kim SM, Eisenberg D: DIP, the database of interacting
780 proteins: a research tool for studying cellular networks of protein interactions. Nucleic Acids Res. 2002, **30**:
781 303-305.
- 782 33. Prasad TSK, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, Telikicherla D, Raju R,
783 Shafreen B, Venugopal A, et al.: Human protein reference database -- 2009 update. Nucleic Acids Res. 2009,
784 **37**: D767-D772.
- 785 34. Kerrien S, Aranda B, Breuza L, Bridge A, Broackes-Carter F, Chen C, Duesbury M, Dumousseau M,
786 Feuermann M, Hinz U, et al.: The IntAct molecular interaction database in 2012. Nucleic Acids Res. 2012,
787 **40**: D841-D846.
- 788 35. Rhee SY, Beavis W, Berardini TZ, Chen G, Dixon D, Doyle A, Garcia-Hernandez M, Huala E, Lander G,
789 Montoya M, et al.: The *Arabidopsis* information resource (TAIR): a model organism database providing a
790 centralized, curated gateway to *Arabidopsis* biology, research materials and community. Nucleic Acids Res.
791 2003, **31**: 224-228.
- 792 36. Lehner B, Fraser AG: A first-draft human protein-interaction map. Genome Biol. 2004, **5**: R63.
- 793 37. Jansen R, Greenbaum D, Gerstein M: Relating whole-genome expression data with protein- protein
794 interactions. Genome Res. 2002, **12**: 37-46.
- 795 38. Udvardi M, Poole PS: Transport and metabolism in legume-rhizobia symbioses. Annu. Rev. Plant Biol. 2013,
796 **64**: 781-805.
- 797 39. Pessi G, Ahrens CH, Rehrauer H, Lindemann A, Hauser F, Fischer HM, Hennecke H: Genome-wide
798 transcript analysis of *Bradyrhizobium japonicum* bacteroids in soybean root nodules. Mol. Plant Microbe
799 Interact. 2007, **20**: 1353-1363.
- 800 40. Cuklina J, Hahn J, Imakaev M, Omasits U, Forstner KU, Ljubimov N, Goebel M, Pessi G, Fischer HM,

- 801 Ahrens CH, et al.: Genome-wide transcription start site mapping of *Bradyrhizobium japonicum* grown
802 free-living or in symbiosis - a rich resource to identify new transcripts, proteins and to study gene regulation.
803 BMC Genomics 2016, **17**: 302.
- 804 41. Delmotte N, Ahrens CH, Knief C, Qeli E, Koch M, Fischer HM, Vorholt JA, Hennecke H, Pessi G: An
805 integrated proteomics and transcriptomics reference data set provides new insights into the *Bradyrhizobium*
806 *japonicum* bacteroid metabolism in soybean root nodules. Proteomics 2010, **10**: 1391-1400.
- 807 42. Grigoriev A: A relationship between gene expression and protein interactions on the proteome scale: analysis
808 of the bacteriophage T7 and the yeast *Saccharomyces cerevisiae*. Nucleic Acids Res. 2001, **29**, 3513-3519.
- 809 43. Ge H, Liu ZH, Church GM, Vidal M: Correlation between transcriptome and interactome mapping data from
810 *Saccharomyces cerevisiae*. Nat. Genet. 2001, **29**: 482-486.
- 811 44. Obayashi T, Kinoshita K: Rank of correlation coefficient as a comparable measure for biological
812 significance of gene coexpression. DNA Research 2009, **16**: 249-260.
- 813 45. von Mering C, Krause R, Snel B, Cornell M, Oliver SG, Fields S, Bork P: Comparative assessment of
814 large-scale data sets of protein-protein interactions. Nature 2002, **417**: 399-403.
- 815 46. Prell J, Poole P: Metabolic changes of rhizobia in legume nodules. Trends Microbiol. 2006, **14**: 161-168.
- 816 47. Nelson MS, Sadowsky MJ: Secretion systems and signal exchange between nitrogen-fixing rhizobia and
817 legumes. Front Plant Sci. 2015, **6**: 491.
- 818 48. Meloni S, Rey L, Sidler S, Imperial J, Ruiz-Argüeso T, Palacios JM: The twin-arginine translocation (Tat)
819 system is essential for *Rhizobium*-legume symbiosis. Molec. Microbiol. 2003, **48**: 1195-1207.
- 820 49. Deakin WJ, Broughton WJ: (2009) Symbiotic use of pathogenic strategies: rhizobial protein secretion
821 systems. Nat. Rev. Microbiol. 2009, **7**: 312-320.
- 822 50. Alloisio N, Queiroux C, Fournier P, Pujic P, Normand P, Vallenet D, Medigue C, Yamaura M, Kakoi K,
823 Kucho K: The *Frankia alni* symbiotic transcriptome. Mol. Plant Microbe Interact. 2010, **23**: 593-607.
- 824 51. Guefrachi I, Pierre O, Timchenko T, Alunni B, Barrière Q, Czernic P, Villaécija-Aguilar JA, Verly C,
825 Bourge M, Fardoux J, et al.: *Bradyrhizobium* BclA is a peptide transporter required for bacterial
826 differentiation in symbiosis with *Aeschynomene* legumes. Mol. Plant Microbe Interact. 2015, **28**: 1155-1166.
- 827 52. Kim DH, Parupalli S, Azam S, Lee SH, Varshney RK: Comparative sequence analysis of nitrogen
828 fixation-related genes in six legumes. Front Plant Sci 2013, **4**: 300.
- 829 53. Jeong H, Mason SP, Barabasi AL, Oltvai ZN: Lethality and centrality in protein networks. Nature 2001, **411**:
830 41-42.
- 831 54. Han JD, Bertin N, Hao T, Goldberg DS, Berriz GF, Zhang LV, Dupuy D, Walhout AJ, Cusick ME, Roth FP,
832 et al.: Evidence for dynamically organized modularity in the yeast protein-protein interaction network.
833 Nature 2004, **430**: 88-93.
- 834 55. Bertolazzi P, Bock ME, Guerra C: On the functional and structural characterization of hubs in protein-
835 protein interaction networks. Biotechnol Adv. 2013, **31**: 274-286.
- 836 56. Kiran M, Nagarajaram HA: Interaction and localization diversities of global and local hubs in human
837 protein-protein interaction networks. Mol. Biosyst. 2016, **12**(9): 2875-2882.
- 838 57. Ungermaun C, Langosch D: Functions of SNAREs in intracellular membrane fusion and lipid bilayer mixing.

- 839 J. Cell Sci. 2005, **118**: 3819-3828.
- 840 58. Hakoyama T, Oi R, Hazuma K, Suga E, Adachi Y, Kobayashi M, Akai R, Sato S, Fukai E, Tabata S, et al.:
- 841 The SNARE protein SYP71 expressed in vascular tissues is involved in symbiotic nitrogen fixation in *Lotus*
- 842 *japonicus* nodules. Plant Physiol. 2012, **160**: 897-905.
- 843 59. Catalano CM, Czymbek KJ, Gann JG, Sherrier DJ: *Medicago truncatula* syntaxin SYP132 defines the
- 844 symbiosome membrane and infection droplet membrane in root nodules. *Planta*, 2007, **225**: 541-550.
- 845 60. Pan H, Oztas O, Zhang X, Wu X, Stonoha C, Wang E, Wang B, Wang D: A symbiotic SNARE protein
- 846 generated by alternative termination of transcription. *Nature Plants* 2016, **2**: 15197.
- 847 61. Li X, Dhaubhadel S: Soybean 14-3-3 gene family: identification and molecular characterization. *Planta* 2011,
- 848 **233**: 569-582.
- 849 62. Radwan O, Wu X, Govindarajulu M, Libault M, Neece DJ, Oh MH, Berg RH, Stacey G, Taylor CG, Huber
- 850 SC, et al.: 14-3-3 proteins SGF14c and SGF14l play critical roles during soybean nodulation. *Plant Physiol.*
- 851 2012, **160**: 2125-2136.
- 852 63. Wienkoop S, Saalbach G: Proteome analysis. Novel proteins identified at the peribacteroid membrane from
- 853 *Lotus japonicus* root nodules. *Plant Physiol.* 2003, **131**: 1080-1090.
- 854 64. Winzer T, Bairl A, Linder M, Linder D, Werner D, Müller P: A novel 53-kDa nodulin of the symbiosome
- 855 membrane of soybean nodules, controlled by *Bradyrhizobium japonicum*. *Mol. Plant Microbe Interact.* 1999,
- 856 **12**: 218-226.
- 857 65. Vinardell JM, Fedorova E, Cebolla A, Kevei Z, Horvath G, Kelemen Z, Tarayre S, Roudier F, Mergaert P,
- 858 Kondorosi A, et al.: Endoreduplication mediated by the anaphase-promoting complex activator CCS52A is
- 859 required for symbiotic cell differentiation in *Medicago truncatula* nodules. *Plant Cell* 2003, **15**: 2093-2105.
- 860 66. Saito K, Yoshikawa M, Yano K, Miwa H, Uchida H, Asamizu E, Sato S, Tabata S, Imaizumi-Anraku H,
- 861 Umehara Y, et al.: NUCLEOPORIN85 is required for calcium spiking, fungal and bacterial symbioses, and
- 862 seed production in *Lotus japonicus*. *Plant Cell* 2007, **19**: 610-624.
- 863 67. Kanamori N, Madsen LH, Radutoiu S, Frantescu M, Quistgaard EM, Miwa H, Downie JA, James EK, Felle
- 864 HH, Haaning LL, et al.: A nucleoporin is required for induction of Ca²⁺ spiking in legume nodule
- 865 development and essential for rhizobial and fungal symbiosis. *Proc. Natl. Acad. Sci. U S A.* 2006, **103**: 359-
- 866 364.
- 867 68. Libault M, Farmer A, Brechenmacher L, Drnevich J, Langley RJ, Bilgin DD, Radwan O, Neece DJ, Clough
- 868 SJ, May GD, et al.: Complete transcriptome of the soybean root hair cell, a single-cell model, and its
- 869 alteration in response to *Bradyrhizobium japonicum* infection. *Plant Physiol.*, 2010, **152**: 541-552.
- 870 69. Sugiyama A, Shitan N, Yazaki K: Involvement of a soybean ATP-binding cassette-type transporter in the
- 871 secretion of genistein, a signal flavonoid in legume-*Rhizobium* symbiosis. *Plant Physiol.* 2007, **144**: 2000-
- 872 2008.
- 873 70. Clarke VC, Loughlin PC, Day DA, Smith PM: Transport processes of the legume symbiosome membrane.
- 874 *Front Plant Sci.* 2014, **5**: 699.
- 875 71. Colebatch G, Desbrosses G, Ott T, Krusell L, Montanari O, Kloska S, Kopka J, Udvardi MK: Global
- 876 changes in transcription orchestrate metabolic differentiation during symbiotic nitrogen fixation in *Lotus*

- 877 *japonicus*. Plant J. 2004, **39**: 487-512.
- 878 72. Limpens E, Bisseling T: Signaling in symbiosis. Curr. Opin. Plant Biol. 2003, **6**: 343-350.
- 879 73. Choudhury SR, Pandey S: Specific subunits of heterotrimeric G proteins play important roles during
880 nodulation in soybean. Plant Physiol., 2013, **162**: 522-533.
- 881 74. Choudhury SR, Pandey S: Phosphorylation-dependent regulation of G-protein cycle during nodule formation
882 in soybean. Plant Cell 2015, **27**: 3260-3276.
- 883 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
884 ROP6 interacts with NFR5 and is involved in nodule formation in *Lotus japonicus*. Plant Physiol. 2012, **159**:
885 131-143.
- 886 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
887 ROP10 of *Medicago truncatula* is required for both tip growth of root hairs and nod factor-induced root hair
888 deformation. Plant Cell 2015, **27**: 806-822.
- 889 77. Harper JF, Harmon A: Plants, symbiosis and parasites: a calcium signalling connection. Nat. Rev. Mol. Cell
890 Biol. 2005, **6**: 555-566.
- 891 78. Bapaume L, Reinhardt D: How membranes shape plant symbioses: signaling and transport in nodulation and
892 arbuscular mycorrhiza. Front Plant Sci. 2012, **3**: 223.
- 893 79. Chen T, Zhu H, Ke D, Cai K, Wang C, Gou H, Hong Z, Zhang Z: A MAP kinase kinase interacts with
894 SymRK and regulates nodule organogenesis in *Lotus japonicus*. Plant Cell 2012, **24**: 823-838.
- 895 80. Stewart GR, Young DB: Heat-shock proteins and the host-pathogen interaction during bacterial infection.
896 Curr. Opin. Immunol. 2004, **16**: 506-510.
- 897 81. Colditz F, Nyamsuren O, Niehaus K, Eubel H, Braun HP, Krajinski F: Proteomic approach: identification of
898 *Medicago truncatula* proteins induced in roots after infection with the pathogenic oomycete *Aphanomyces*
899 *euteiches*. Plant Mol. Biol. 2004, **55**: 109-120.
- 900 82. Oehrle NW, Sarma AD, Waters JK, Emerich DW: Proteomic analysis of soybean nodule cytosol.
901 Phytochemistry 2008, **69**: 2426-2438.
- 902 83. Salavati A, Taleei A, Bushehri AA, Komatsu S: Analysis of the proteome of common bean (*Phaseolus*
903 *vulgaris* L.) roots after inoculation with *Rhizobium etli*. Protein & Peptide Letters 2012, **19**: 880.
- 904 84. Panter S, Thomson R, de Bruxelles G, Laver D, Trevaskis B, Udvardi M: Identification with proteomics of
905 novel proteins associated with the peribacteroid membrane of soybean root nodules. Mol. Plant Microbe
906 Interact. 2000, **13**: 325-333.
- 907 85. Catalano CM, Lane WS, Sherrier DJ: Biochemical characterization of symbiosome membrane proteins from
908 *Medicago truncatula* root nodules. Electrophoresis 2004, **25**: 519-531.
- 909 86. Ashby MK: Survey of the number of two-component response regulator genes in the complete and annotated
910 genome sequences of prokaryotes. FEMS Microbiol. Lett. 2004, **231**: 277-281.
- 911 87. Whitworth DE, Cock PJ: Evolution of prokaryotic two-component systems: insights from comparative
912 genomics. Amino Acids 2009, **37**: 459-466.
- 913 88. Charles TC, Jin S, Nester EW: Two-component sensory transduction systems in phytochromata. Annu. Rev.
914 Phytopathol. 1992, **30**: 463-484.
- 915 89. West AH, Stock AM: Histidine kinases and response regulator proteins in two-component signaling systems.

- 916 Trends Biochem. Sci. 2001, **26**: 369-376.
- 917 90. Hunter WJ: Increased nodulation of soybean by a strain of *Bradyrhizobium japonicum* with altered
918 tryptophan metabolism. Lett. Appl. Microbiol. 1994, **18**: 340-342.
- 919 91. Lambrecht M, Okon Y, Vande Broek A, Vanderleyden J: Indole-3-acetic acid: a reciprocal signalling
920 molecule in bacteria-plant interactions. Trends Microbiol. 2000, **8**: 298-300.
- 921 92. Ghosh S, Basu PS: Production and metabolism of indole acetic acid in roots and root nodules of *Phaseolus*
922 *mungo*. Microbiol. Res. 2006, **161**: 362-366.
- 923 93. Roux B, Rodde N, Jardinaud MF, Timmers T, Sauviac L, Cottret L, Carrère S, Sallet E, Courcelle E, Moreau
924 S, et al.: An integrated analysis of plant and bacterial gene expression in symbiotic root nodules using
925 laser-capture microdissection coupled to RNA sequencing. Plant J. 2014, **77**: 817- 837.
- 926 94. Sokolovski S, Hills A, Gay RA, Blatt MR: Functional interaction of the SNARE protein NtSyp121 in Ca²⁺
927 channel gating, Ca²⁺ transients and ABA signalling of stomatal guard cells. Mol. Plant 2008, **1**: 347-358.
- 928 95. Lodwig E, Poole P: Metabolism of *Rhizobium* bacteroids. Crit. Rev. Plant Sci. 2003, **22**, 37-78.
- 929 96. Oke V, Long SR: Bacteroid formation in the rhizobium-legume symbiosis. Curr. Opin. Microbiol. 1999, **2**,
930 641-646.
- 931 97. Kersey PJ, Allen JE, Armean I, Boddu S, Bolt BJ, Carvalho-Silva D, Christensen M, Davis P, Falin LJ,
932 Grabmueller C, et al.: Ensembl Genomes 2016: more genomes, more complexity. Nucleic Acids Res. 2016,
933 **44**: D574-D580.
- 934 98. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N,
935 et al.: Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res. 2012, **40**:
936 D1178-D1186.
- 937 99. Li QG, Zhang L, Li C, Dunwell JM, Zhang YM: Comparative genomics suggests that an ancestral
938 polyploidy event leads to enhanced root nodule symbiosis in the Papilionoideae. Mol. Biol. Evol. 2013, **30**:
939 2602-2611.
- 940 100. Sonnhammer EL, Östlund G: InParanoid 8: orthology analysis between 273 proteomes, mostly eukaryotic.
941 Nucleic Acids Res. 2015, **43**: D234-D239.
- 942 101. Yellaboina S, Tasneem A, Zaykin DV, Raghavachari B, Jothi R: DOMINE: a comprehensive collection of
943 known and predicted domain-domain interactions. Nucleic Acids Res. 2011, **39**: D730-D735.
- 944 102. Finn RD, Coghill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M,
945 Sangrador-Vegas A, et al.: The Pfam protein families database: towards a more sustainable future. Nucleic
946 Acids Res. 2016, **44**: D279-D285.
- 947 103. O'Brien KP, Westerlund I, Sonnhammer EL: OrthoDisease: a database of human disease orthologs. Hum.
948 Mutat. 2004, **24**, 112-119.
- 949 104. O'Brien KP, Remm M, Sonnhammer EL: Inparanoid: a comprehensive database of eukaryotic orthologs.
950 Nucleic Acids Res. 2005, **33**: D476-D480.
- 951 105. Krogh A, Larsson B, von Heijne G, Sonnhammer EL: Predicting transmembrane protein topology with a
952 hidden Markov model: application to complete genomes. J. Mol. Biol. 2001, **305**: 567-580.
- 953 106. 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, D é nari é 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, Daskocilova 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, Sch ütze K, Batistic O, Weckermann K, N ä ke 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

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
<i>Bradyrhizobium diazoefficiens</i> 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
<i>Glycine max</i>					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					

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

<i>Glycine max</i>					<i>Bradyrhizobium diazoefficiens</i> 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	bj00190	20	66	1.46E-10
Phagosome	gmx04145	52	165	7.00E-07	Protein export	bj03060	10	20	5.58E-07
Protein export	gmx03060	36	91	8.57E-07	Two-component system	bj02020	20	168	5.79E-05
Protein processing in endoplasmic reticulum	gmx04141	81	375	6.27E-05	Peptidoglycan biosynthesis	bj00550	6	24	0.003838
N-Glycan biosynthesis	gmx00510	27	74	9.75E-05	Glycerophospholipid metabolism	bj00564	6	24	0.003838
Ribosome	gmx03010	109	595	0.000594	ABC transporters	bj02010	22	299	0.007615
Biosynthesis of amino acids	gmx01230	77	429	0.011763	Bacterial secretion system	bj03070	7	44	0.009047
Citrate cycle (TCA cycle)	gmx00020	26	104	0.015133	beta-Lactam resistance	bj01501	5	21	0.009047
Carbon metabolism	gmx01200	84	488	0.016097					
Proteasome	gmx03050	24	105	0.049615					

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

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

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

989 $\text{sim}_{\text{JC}}^{\text{CC}}$; C: distribution of $\text{sim}_{\text{JC}}^{\text{MF}}$. 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

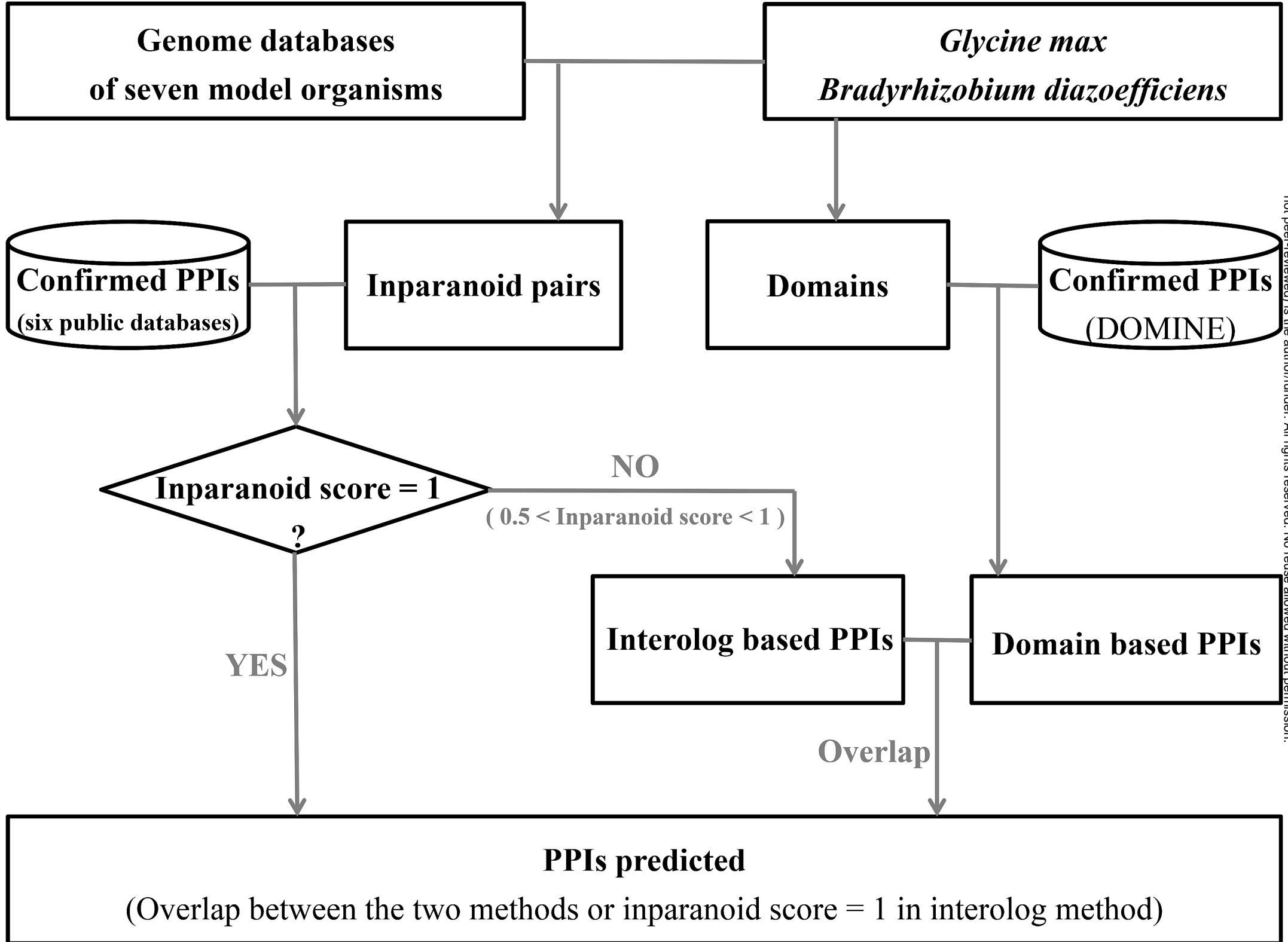
1002 half of LUC were fused to N-gene, N-LUC, C-gene and C-LUC. In (c), the treatment

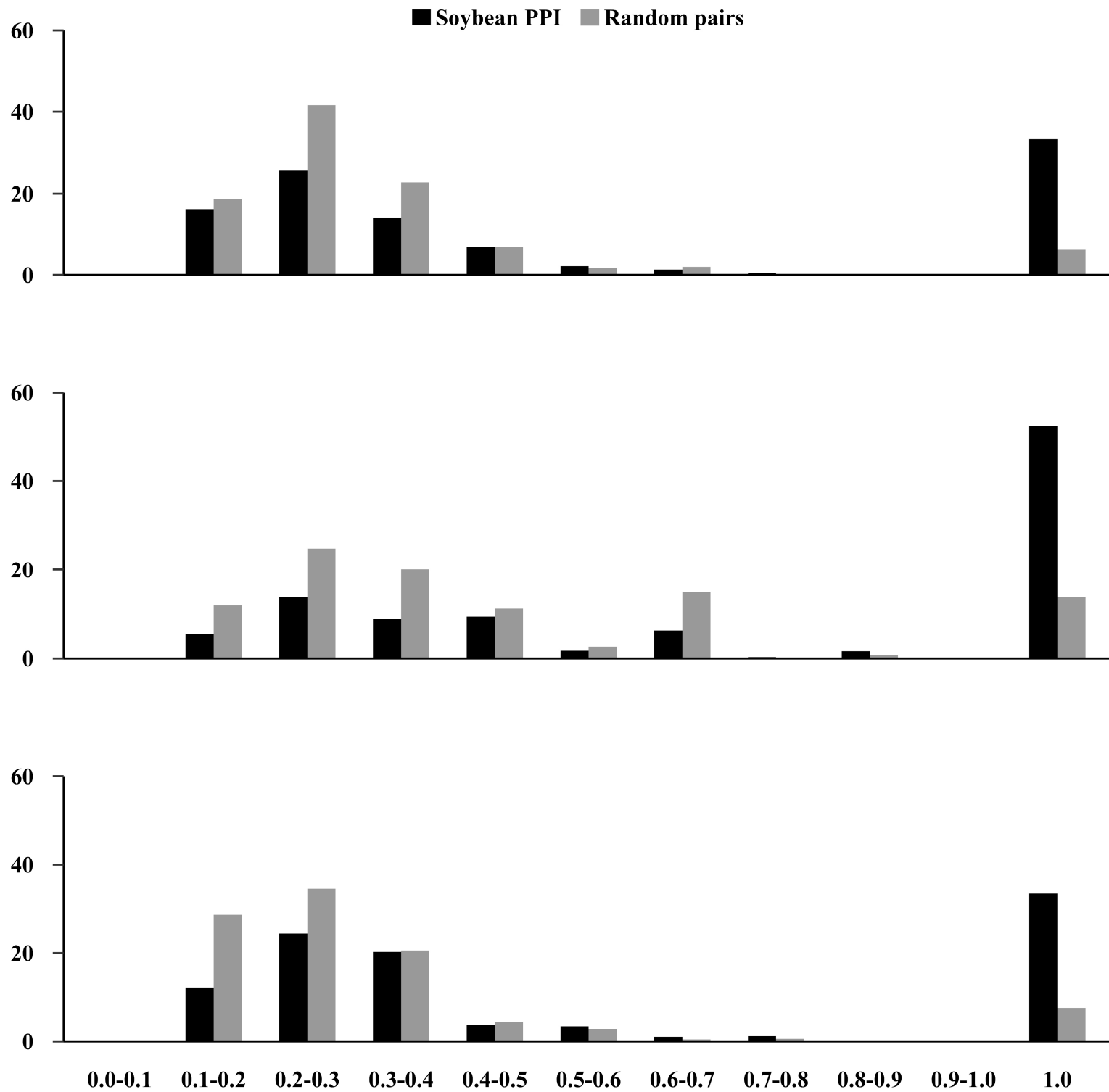
1003 was N-GmSGF14g + C-BAC48988, and the controls were N-LUC + C-BAC48988,

1004 N-GmSGF14g + C-LUC, and N-LUC + C-LUC. LUC fluorescence was detected by

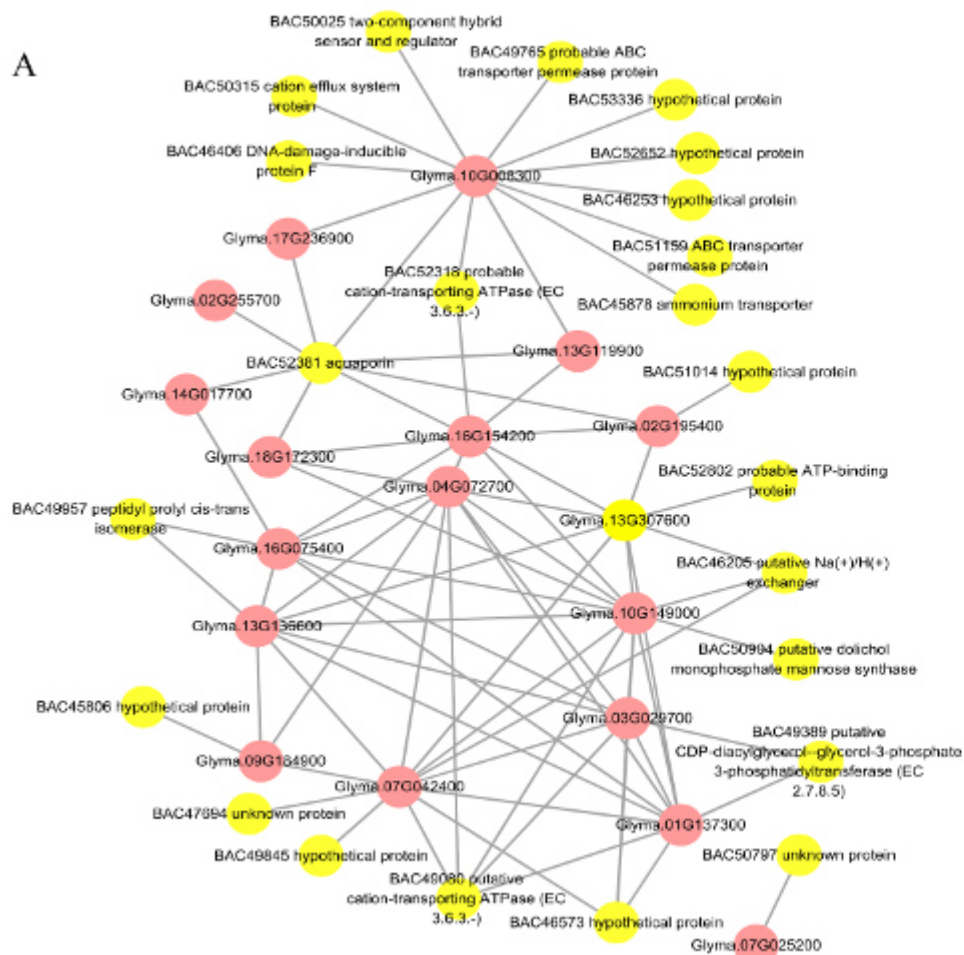
1005 confocal microscope in *N. benthamiana* fresh leaves. The experiment was repeated

1006 three times with similar results. The situation was similar in the others.





A



B

