Article

Computationally Reconstructed Interactome of *Bradyrhizobium diazoefficiens* USDA110 Reveals Novel Functional Modules and Protein Hubs for Symbiotic Nitrogen Fixation

Jun-Xiao Ma [†], Yi Yang [†], Guang Li (ID) **and Bin-Guang Ma ***

Hubei Key Laboratory of Agricultural Bioinformatics, College of Informatics, Huazhong Agricultural University, Wuhan 430070, China; majunxiao0518@163.com (J.-X.M.); yyphoenix@163.com (Y.Y.); rwwb@foxmail.com (G.L.)
* Correspondence: mbg@mail.hzau.edu.cn; Fax: +86-27-8728-0877
† These authors contributed equally to this work.

Abstract: Symbiotic nitrogen fixation is an important part of the nitrogen biogeochemical cycles and the main nitrogen source of the biosphere. As a classical model system for symbiotic nitrogen fixation, rhizobium-legume systems have been studied elaborately for decades. Details about the molecular mechanisms of the communication and coordination between rhizobia and host plants is becoming clearer. For more systematic insights, there is an increasing demand for new studies integrating multiomics information. Here, we present a comprehensive computational framework integrating the reconstructed protein interactome of *B. diazoefficiens* USDA110 with its transcriptome and proteome data to study the complex protein-protein interaction (PPI) network involved in the symbiosis system. We reconstructed the interactome of *B. diazoefficiens* USDA110 by computational approaches. Based on the comparison of interactomes between *B. diazoefficiens* USDA110 and other rhizobia, we inferred that the slow growth of *B. diazoefficiens* USDA110 may be due to the requirement of more protein modifications, and we further identified 36 conserved functional PPI modules. Integrated with transcriptome and proteome data, interactomes representing free-living cell and symbiotic nitrogen-fixing (SNF) bacteroid were obtained. Based on the SNF interactome, a core-sub-PPI-network for symbiotic nitrogen fixation was determined and nine novel functional modules and eleven key protein hubs playing key roles in symbiosis were identified. The reconstructed interactome of *B. diazoefficiens* USDA110 may serve as a valuable reference for studying the mechanism underlying the SNF system of rhizobia and legumes.

Keywords: rhizobia; protein-protein interaction; network analysis; multiomics; data integration; symbiosis

1. Introduction

In most cases, rhizobia are a group of Gram-negative soil bacteria within the family *Rhizobiaceae*. They can colonize roots of legumes, establish symbiotic relationship with them and perform nitrogen fixation. This kind of symbiotic nitrogen-fixing (SNF) system can convert inorganic nitrogen to organic nitrogen and constitutes an important part of the biogeochemical cycle [1]. *B. diazoefficiens* USDA110, (formerly known as *B. japonicum* USDA110) [2], a strain of *Bradyrhizobium*, which is an important genus of rhizobia, can establish a symbiotic relationship with *Glycine max* (soybean). The *G. max*–*B. diazoefficiens* system has been an important model for the study of the SNF system [3,4]. In addition, given the efficient SNF ability of *B. diazoefficiens* USDA110 [5], it is widely used in agriculture and environmental engineering [6,7]. Owing to its importance, various biological characteristics of *B. diazoefficiens* have been extensively studied for decades, especially its SNF mechanism, such as the information exchange between *B. diazoefficiens* and *G. max* during the process

of establishing symbiosis [8,9]. The genome of *B. diazoefficiens* reference strain USDA110 was completely sequenced in 2002 [10], and corresponding high-throughput omics studies were performed in recent years [11–17]. Although a large amount of knowledge has been obtained through these studies, contribution of the protein-protein interaction (PPI) network of this species in the SNF mechanism remains elusive.

The proteins in cells do not exist in isolation. Biological processes often involve multiple proteins and their interactions [18,19]. Thousands of proteins are coordinated, forming a functional network, the PPI network, which is becoming one of the main research objects in systems biology. PPI networks have become effective tools for understanding cellular behaviors and solving various biological problems in transcription, translation, metabolism, gene regulation and signal transduction [20]. From the more comprehensive perspective of a PPI network, new functions of proteins have been discovered and new insights gained [21–23]. In addition, based on the information of proteins with precise annotations and the relationships between proteins in the PPI networks, the functions of poorly functional characterized proteins can be inferred and knowledge gaps can be filled [24,25].

Several high-throughput experimental techniques, such as yeast two-hybrid (Y2H), Tandem Affinity Purification and Mass Spectrometry (TAP-MS), and protein chips, have made significant contributions to the detection of PPIs for constructing genome-scale PPI networks [26–28]. However, these methods are intensive in labor and funds, let alone their inherent biases and limited coverage [29]. As far as we know, the number of PPIs detected by experimental methods are far from the estimated amount [30]. As a result, several types of in silico methods have been developed to meet the demand for determining missing PPIs in more species [31–33]. In this study, the protein interactome of *B. diazoefficiens* USDA110 was inferred by using "Interolog" and domain-based methods and reconstructed into a PPI network. The network was analyzed from different perspectives and compared with the PPI networks of other rhizobia. By integrating transcriptome and proteome data, PPI networks representing two typical physiological states-free-living (FL) cell and SNF bacteroid-were obtained and compared. Based on the SNF network, a core-sub-PPI-network related to symbiotic nitrogen fixation was determined and dissected, and novel functional modules and protein hubs were identified.

2. Materials and Methods

2.1. Generation of Protein Interactome

2.1.1. Interolog Method

An "interolog" is a conserved interaction between a pair of proteins which have interacting homologs in another organism [34,35]. For two proteins in *B. diazoefficiens* USDA110, if their homologous proteins in other species have an interaction, they are predicted to be interacting. For the "Interolog" method, we selected six prokaryotes as primary reference species including *Campylobacter jejuni*, *Escherichia coli*, *Helicobacter pylori*, *Mesorhizobium loti*, *Synechocystis sp.* PCC6803 and *Treponema pallidum* (Figure 1), all of which have relatively well-developed information of PPIs determined by experiments. PPIs of the six reference species were collected from the literature [36–41]. All protein sequences of *B. diazoefficiens* USDA110 and of the six reference species were downloaded from the NCBI RefSeq database. We predicted PPIs of *B. diazoefficiens* USDA110 by the "Interolog" method in two ways. On one hand, protein sequences of the six reference species were prepared to do homologous alignment with *B. diazoefficiens* USDA110, and on the other hand, experimental PPIs and their protein sequences from BioGRID [42], DIP [43], IntAct [44] and HPRD_Release9_062910 [45] databases were added as supplements regardless of species. In order to improve the prediction accuracy, only the protein sequences from those databases which also appeared in the Swiss-Prot database were used. The homologous proteins were acquired by using InParanoid [46] (version 4.1) under default parameters.
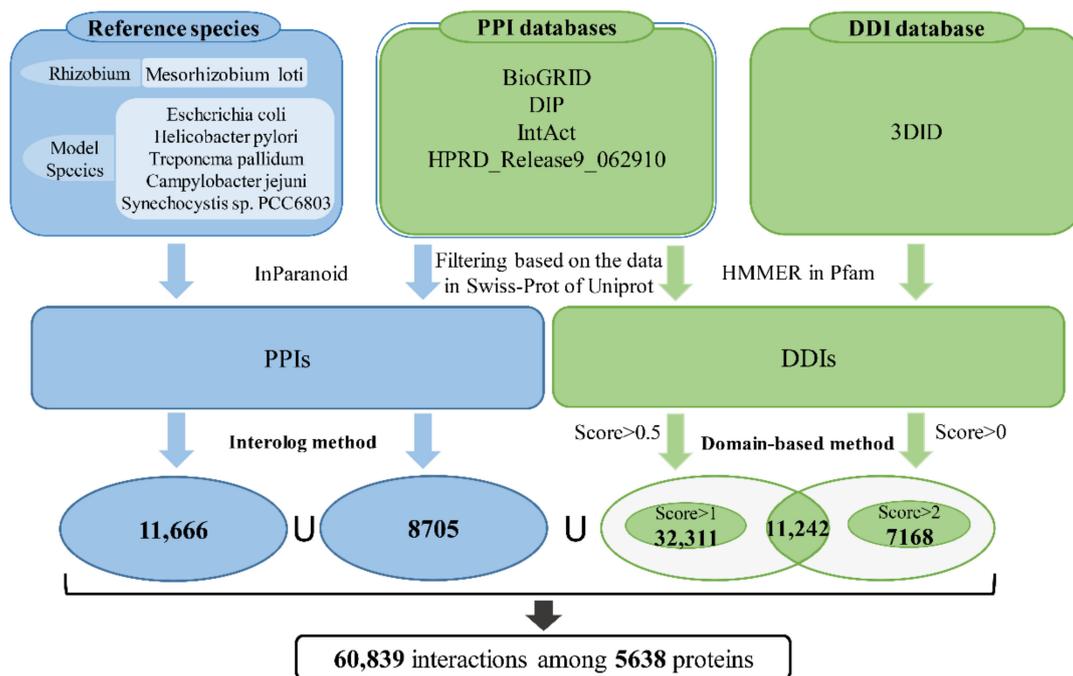
Figure 1. Flowchart for reconstructing the protein interactome of *B. diazoefficiens* USDA110.

2.1.2. Domain-Based Method

In recent years, the domain-based method has been widely used in PPI prediction because the occurrence of PPI is usually caused by the interaction between one domain X of protein A and another domain Y of protein B, and the interaction between proteins A and B can be inferred if domains X and Y have verified interaction. Here, domain-domain interactions (DDIs) were obtained at both the sequence and structure levels. The PPIs in databases of BioGRID, DIP, IntAct and HPRD_Release9_062910 (Figure 1) were utilized to infer DDIs. Besides, DDIs were also acquired from the three-dimensional interacting domains (3DID) database [47]. Protein domains in *B. diazoefficiens* USDA110 and in BioGRID, DIP, IntAct and HPRD_Release9_062910 databases were recognized and obtained based on the domain definitions of the Pfam database and HMMER program (score > 20, *e*-value < 1×10^{-5}, coverage > 0.9) [48,49]. A part of DDIs were inferred from experimental PPIs in databases based on domain sequences with score > 0.5, and another part of DDIs were obtained from 3DID database based on the domain structure with score > 0. For the former part, if the proteins in a PPI pair were regarded to have domains based on protein sequences (score > 0.5) and the PPI had relevant experimental verification in at least one of these databases (BioGRID, DIP, IntAct and HPRD), the corresponding DDI was inferred. As shown in Figure 1, the intersection of these two parts, as well as the domains of score > 1 from PPI databases, and the domains of score > 2 from 3DID database, were used in order to improve accuracy and data coverage.

2.2. Quality Assessment of the Reconstructed PPI Network

First, the PPIs were verified by an iLoop server (http://sbi.imim.es/iLoops.php, 5 May 2017) which utilizes local structure features to characterize protein interactions [50,51]. In our study, 1000 PPIs were randomly selected and submitted to the iLoop server for validation. Second, considering the interacting proteins are likely to appear in the same subcellular localization, the PPIs were verified by subcellular colocalization of the proteins. Sequences of the 5638 proteins in the *B. diazoefficiens* USDA110 interactome were submitted to PSORT 3.0 (http://www.psort.org/psortb, 16 May 2017) [52] and the location information of each protein was obtained. Third, the two proteins in a PPI tend to have similar functions. Based on this point, the functional similarity of each interacting protein pair in the reconstructed PPI network, and the randomized PPI network of the same topology,

were calculated according to a published algorithm G-SESAME [53]. The Gene Ontology (GO) function annotations of *B. diazoefficiens* USDA110 proteins were obtained from the GO database [54], and the Wilcoxon rank-sum test was performed to compare the functional similarity difference between the reconstructed PPI network and the random PPI network. Last, the correlation of transcriptional profiles between two genes from a PPI pair was also used to examine the reliability of the network. Here, the SNF subnetwork of the global network was checked based on time-sequenced transcriptome data [11]. The Pearson Correlation Coefficient (PCC) was calculated for the transcription profiles of every interacting pair in the reconstructed PPI network and the randomized PPI network of the same topology, and then Wilcoxon rank-sum test was used to show the difference. All of the methods used to verify the interactions are yes-no methods, meaning that they only give information on whether or not the two proteins have an interaction.

2.3. COG Function Enrichment Analysis of PPIs

We assorted PPI numbers according to the Clusters of Orthologous Groups of proteins (COG) annotations as the functional network. Referring to 1000 random networks of the same topology generated from the functional network, the COG functional network was presented in the form of heat map. Z-scores were calculated to determine colors in the heat map by the following formula:

$$Z_{ij} = \frac{A_{ij} - <Rnd_{ij}>}{\sigma_{Rnd_{ij}}}$$

where A_{ij} represents the actual number of interacting protein pairs between COG category i and category j, and $<Rnd_{ij}>$ and $\sigma_{Rnd_{ij}}$ represent the mean and standard deviation of interacting protein pairs between category i and category j summarized from the 1000 random networks, respectively.

2.4. Reconstruction and Validation of the PPI Sub-Networks in FL and SNF Conditions

B. diazoefficiens USDA110 cells cultivated in aerobic peptone salts-yeast extract medium were considered as the FL condition. The processed microarray data of the FL *B. diazoefficiens* USDA110 were downloaded from NCBI-GEO database [55] (GSM210269-GSM210283). In order to reduce false positives, a gene was deemed as expressed only when more than 80% of the replicates were of p-value ≤ 0.06 [11]. Then, the average of the signals in selected replicates was regarded as the expression level [11]. The transcriptome and proteome data of bacteroids under the SNF state were obtained directly from the Supplementary Materials of a previous study [14]. The FL and SNF subnetworks were reconstructed by selecting PPIs whose components were expressed under the FL and SNF conditions, respectively, indicated by transcriptome and proteome data [11,14]. The transcription level difference of an interacting protein pair i and j was normalized according to the following formula [56]:

$$D_{ij} = \frac{|Signal_i - Signal_j|}{|Signal_i + Signal_j|}$$

where $Signal_i$ and $Signal_j$ are the transcription levels of gene i and gene j, respectively.

3. Results and Discussion

3.1. Reconstruction of the Genome-Scale PPI Network of B. diazoefficiens USDA110

The PPIs in *B. diazoefficiens* USDA110 were predicted by integration of an "Interolog" and a domain-based method (Figure 1). The reconstructed PPI network was visualized by using Cytoscape 3.2.1 [57] (Figure S1A), which contains 5638 proteins and 60,839 PPIs. The Clusters of Orthologous Groups of proteins (COG) functional categories [58] of the nodes (proteins) in the PPI network were determined, and their proportions are shown in Figure S1B. Proteins related to 'amino acid transport and metabolism (E)' accounted for the

largest proportion (over 12%), while the proteins associated with 'cell cycle control, cell division, chromosome partitioning (D)' only accounted for 0.5%.

3.2. Quality Assessment of the PPI Network

Reliability of the reconstructed PPI network was assessed from four different perspectives: local structural features, subcellular localization, functional similarities and transcriptional correlations of genes. For local structural features, 1000 PPIs were selected randomly and submitted to the iLoop server by which the interactions were predicted. Due to the limitation of available structural templates (in the server's dependent databases), 47.4% of the 1000 PPIs had at least one protein without corresponding structural features and thus could not be assessed. For the remaining PPIs, 46.4% were confirmed, whereas only 1% were identified as noninteracting pairs (Figure 2A), which proved that the reconstructed PPI network is reliable based on local structural features.
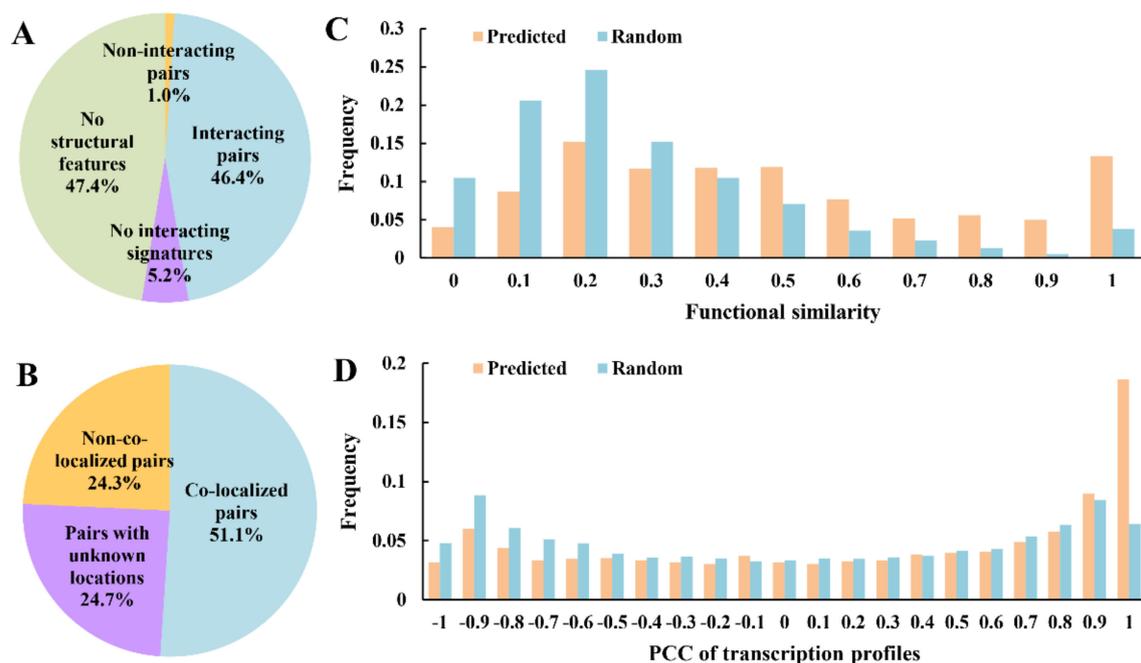

Figure 2. Validation of the *B. diazoefficiens* USDA110 protein interactome. (**A**) Validated by iLoop server. "Interacting pairs" represents the directly confirmed PPIs; "Non-interacting pairs" represents the directly non-confirmed PPIs; "No structural features" represents that no structural features (loop or domain) could be assigned to either one or both proteins; "No interaction signatures" represents that the pair of proteins had structural features but no interacting features. (**B**) Validated by subcellular co-localization. "Co-localized pairs" represents that the two interacting proteins were co-localized; "Non-co-localized pairs" represents that the two interacting proteins were not co-localized; "Pairs with unknown locations" represents that at least one protein in an interacting protein pair had no subcellular location information. (**C**) The difference of distribution of functional similarity between the reconstructed network and the randomized network of the same topology. Higher frequencies appear in the reconstructed network than the random network when functional similarity becomes larger. (**D**) The distribution difference of PCC of gene transcription profiles between the reconstructed network and the randomized network of the same topology. Higher frequencies appear in the reconstructed network than the random network when PCC of transcription profiles becomes larger.

Proteins that interact with each other are likely to appear in the same subcellular location. Here, we ascertained the subcellular location of the 5638 proteins in the reconstructed PPI network. According to the results, more than half of the PPIs (51.1%) were confirmed to be colocalized in the cell. Besides, over 25% PPI pairs had a component whose location was unknown (Figure 2B), and for these PPI pairs some of them must be colocalized; therefore, the proportion of positive results was actually larger than 51.1%.

Previous studies have demonstrated that interacting proteins tend to have relatively higher functional similarities [38,59], which can be used to confirm the reliability of the reconstructed PPI network. Based on the semantic similarities of gene ontology (GO) annotations [54], the functional similarities of the proteins in PPI pairs of the reconstructed PPI network and corresponding random networks of the same topology were calculated and compared (Figure 2C). This showed that the functional similarities in the reconstructed PPI network were significantly (p-value $< 2.2 \times 10^{-16}$ from Wilcoxon rank-sum test) higher than those of the random networks, which proved that the reconstructed PPI network is reliable based on functional similarities of the proteins in PPI pairs.

It has been reported that interacting proteins have similar expression patterns [60]. Based on the temporal transcriptome profiles of the bacteroid of *B. diazoefficiens* USDA110 [11], the PCC of normalized transcription data of the proteins in PPI pairs of the reconstructed PPI network, and corresponding random networks with the same topology, were calculated and compared (Figure 2D). The PCC values of the reconstructed PPI network were significantly (p-value $< 2.2 \times 10^{-16}$ from Wilcoxon rank-sum test) higher than those of the random networks. When PCC is close to 1.0, the difference is more significant, which means that the reconstructed PPI network contained much more PPI pairs whose components kept fairly consistent transcriptional patterns than random networks. Hence, the reconstructed PPI network is reliable based on coexpression of genes.

3.3. Properties of the Reconstructed PPI Network

The topological parameters of the PPI network were calculated and analyzed with the "Network Analysis" plugin in Cytoscape [61]. Degree distribution of the PPI network of *B. diazoefficiens* USDA110 follows the power law ($y = 2143.3x^{-1.348}$, $R^2 = 0.76$, Figure 3A), which shows the characteristic of scale-free network, such as many typical biological networks [62]. The number of subnetworks is 46 (Figure S1A), and the largest sub-network contains 5556 proteins and 60,730 interactions. The node degree distribution is shown in Figure 3B. The clustering coefficient is always used to evaluate the modularity of a node in the PPI network by examining the connectivity among the neighbors of the current node. A higher clustering coefficient suggests that the protein is more likely to participate in a close-connected module. The distribution of the clustering coefficient of the PPI network had two peaks (Figure 3C), suggesting that the topological parameters of series of small sub-networks were very different from those of the largest one [63]. The average shortest path length was around 3 and 4 (Figure 3D), which is consistent with the results of previous studies [64].

In the power law fitting formula, the degree exponent γ of the reconstructed PPI network was calculated as 1.348 by the maximum likelihood estimate. Studies have shown that if the degree exponent is smaller than 2, the control of the whole network requires a small set of essential nodes, which could be identified by minimum dominating set (MDS) [23,65]. By solving an integer-based linear programming problem [65], a MDS of the *B. diazoefficiens* USDA110 PPI network was determined. In order to eliminate the bias caused by self-interactions, they were removed from the reconstructed PPI network, and the refined network (including 5611 proteins and 59,125 PPIs) was used to identify the MDS. The determined MDS of the PPI network contained 427 nodes (less than 10% of the total number). Furthermore, COG enrichment analysis was performed to examine the functional distribution of these essential nodes. As shown in Table S1, the proteins in MDS are significantly enriched (Fisher's exact test, $p < 0.01$) in 'Energy production and conversion (C)', 'Lipid transport and metabolism (I)', 'Posttranslational modification, protein turnover, chaperones (O)', 'Intracellular trafficking, secretion, and vesicular transport (U)' and 'Defense mechanisms (V)'.
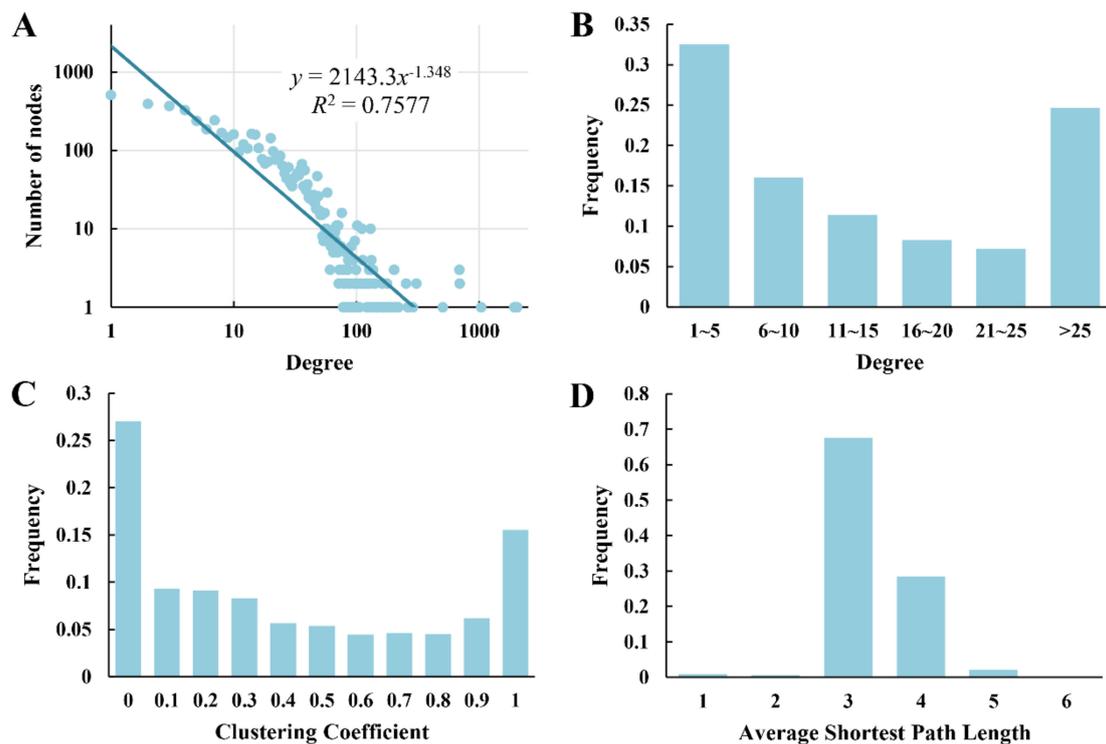
Figure 3. Distributions of topological properties of the *B. diazoefficiens* USDA110 PPI network.

3.4. Comparison of PPIs among Rhizobia

In the reconstructed PPI network of *B. diazoefficiens* USDA110, 5199 proteins with COG annotations (involving 55,668 PPIs) were selected and analyzed. A z-score index, which quantitatively reflects the degree of enrichment of PPIs in different combinations of COG categories, was defined and calculated [41,66]. Same calculations were performed on the PPI networks of *M. loti* (2377 PPIs among 1408 proteins) [39], *S. meliloti* (856 PPIs among 320 proteins) [67] and *B. diazoefficiens* USDA110 under the SNF state (10,473 PPIs among 1777 proteins). As shown in Figure 4, proteins of the same functional categories are more likely to interact with each other in *B. diazoefficiens* USDA110 and *S. meliloti*, while this feature is not obvious in *M. loti*. Moreover, the PPIs related to the proteins of category O (post translational modification, protein turnover, chaperones) are the mainstay in the PPI network of *B. diazoefficiens* USDA110, because this protein category interacts with almost all the other protein categories. In contrast, the enrichment of PPIs related to category O in *M. loti* and *S. meliloti* is not obvious. This result may indicate that the PPIs in rhizobia are correlated with growth rate, and the slow growth of *B. diazoefficiens* USDA110 requires more protein modification and repair that are performed by the proteins in category O. Another noticeable point is the interaction between protein categories J (translation, ribosomal structure and biogenesis) and U (intracellular trafficking, secretion, and vesicular transport), enriched in the *B. diazoefficiens* (SNF) and *S. meliloti* PPI networks, which are both subnetworks for symbiotic nitrogen fixation, but not so in *B. diazoefficiens* and *M. loti*, which are both global PPI networks. This result may suggest that protein biogenesis and cellular material transportation are correlated with each other through PPI in the SNF state in rhizobia. This enhances our understanding on the roles of material supply in the mechanism of symbiotic nitrogen fixation [68].
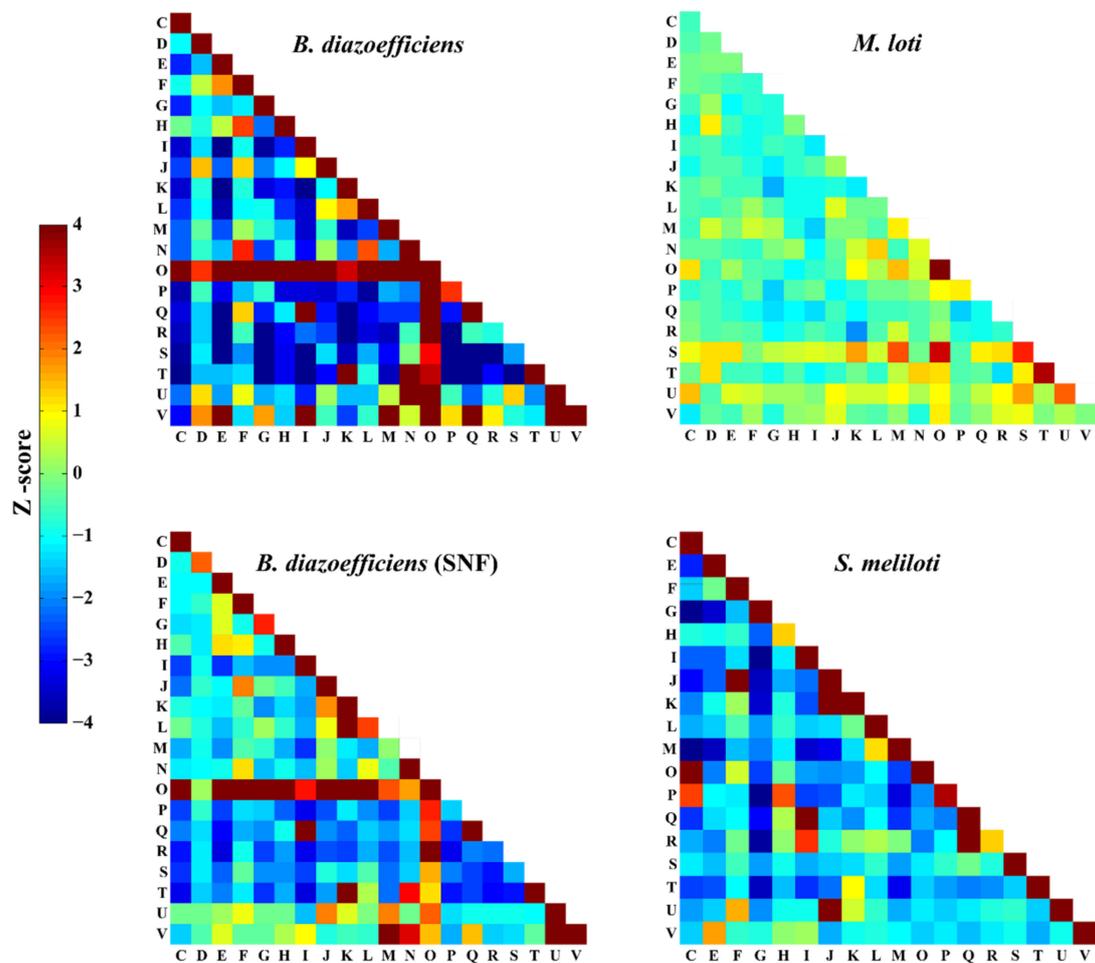
Figure 4. Enrichment in COG functional categories of PPIs for three organisms in different situations (the global PPI network: *B. diazoefficiens* USDA110, *M. loti*; the PPI network for SNF: *B. diazoefficiens* USDA110 (SNF), *S. meliloti*). The PPI numbers in COG categories were normalized by Z-score. Warmer color means higher enrichment.

Common functional modules, such as signal pathways and protein complexes, are relatively conservative among different species [69,70]. An algorithm called GASOLINE is available for the comparison of biological networks to find conserved functional modules [71]. We used this algorithm to compare the PPI networks of several rhizobia for both the global network and the subnetwork under the SNF state. The global PPI networks of *B. diazoefficiens* USDA110 and *M. loti* had 26 pairs of conserved functional modules (Density threshold = 0.5, Sigma = 2) with a particularly high-quality score (Table S2). For example, the largest module with the highest score (module 25 in Table S2, six proteins with ISC score 0.81) included the hybrid sensors and regulators of the two-component system for chemotaxis. The protein interactions within this module and the correspondence relationships between the two compared species are shown in Figure 5A. Since these species have special requirements for swimming in soil [72], this module is considered to play a regulatory role in flagellum movement when rhizobia meet attractants in rhizosphere. As shown in Figure 5A, eight PPIs in this module match well between *B. diazoefficiens* USDA110 and *M. loti*. Considering the two additional PPIs (blr2194-bll1199, blr2192-bll1199) in *B. diazoefficiens* USDA110 and the correspondence relationship, it can be hypothesized that Q98HL7-Q98L88 and Q98PD0-Q98L88 in *M. loti* are also interacting protein pairs that might have been missed in the experiments in [39].
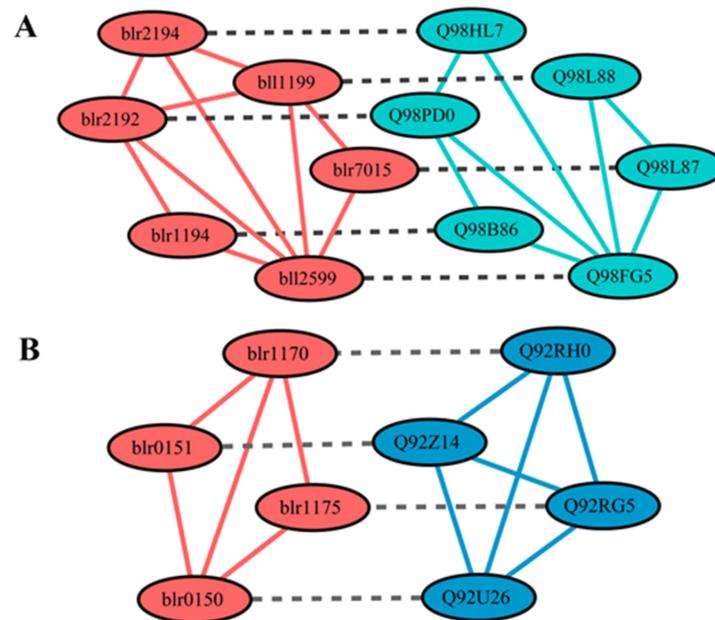
Figure 5. Examples of conserved PPI modules identified by GASOLINE. (**A**) The conserved PPI module related to chemotaxis was identified from the comparison between the global PPI network of *B. diazoefficiens* USDA110 (red) and *M. loti* (green). (**B**) The conserved PPI module related to cytochrome C oxidase was identified from the comparison between the SNF PPI network of *B. diazoefficiens* USDA110 (red) and *S. meliloti* (blue). Solid lines represent the PPI within each module and dashed lines represent correspondence (homology) relationship between species.

For the SNF state, 10 pairs of conserved functional modules were obtained (Density threshold = 0.7, Sigma = 3) by comparing the PPI networks of *B. diazoefficiens* USDA110 and *S. meliloti* (Table S3). For example, the module with the highest score (module 1 in Table S3, 4 proteins with ISC score 0.88) includes ubiquinol-cytochrome enzyme (complex III) subunit and cytochrome c oxidase (complex IV) subunit. The protein interactions within this module and the correspondence relationships between the two species are shown in Figure 5B. Since nitrogen fixation usually requires additional energy in terms of ATP and reductant, cytochrome c oxidase may produce the required ATPs to provide the extra energy consumption [73]. Bhargava et al. proved that the activity of cytochrome c oxidase was enhanced in the process of nitrogen fixation [74]. The existence of the above PPI module (Figure 5B), identified from *B. diazoefficiens* USDA110 and *S. meliloti,* further validates the importance of cytochrome c oxidase in nitrogen fixation. Previous studies have shown that the process of electron transfer between complex III and complex IV needs cytochrome c, but the specific process is still unknown [75]. In our study, the subunit components of complex III (blr0151, blr0150; Q92Z14, Q92U26) and the subunit components of complex IV (blr1170, blr1175; Q92RH0, Q92RG5) all have interactions (Figure 5B). Furthermore, in the 3DID database [47], the domains (cox1, cox2, cox3) of subunits in complex III and complex IV were identified to be interacting. Based on this evidence, it can be inferred that complex III and complex IV may have direct interaction under some states to improve the efficiency of electron transfer and release more energy for nitrogen fixation.

In addition, with the reconstructed network of *B. diazoefficiens* USDA110, by network comparison with *M. loti* using GASOLINE and by consulting GO and KEGG databases, the functions of proteins belonging to COG category S ("function unknown") were inferred (Table S4), which is a direct application of our reconstructed PPI network of *B. diazoefficiens* USDA110.

3.5. Analysis and Comparison of the PPI Networks under FL and SNF States

As the type strain of *Bradyrhizobium* rhizobia, *B. diazoefficiens* USDA110 has two typical physiological states: the FL cell and the SNF bacteroid. Therefore, it is meaningful to investigate the differences and similarities of the reconstructed PPI networks in these two states. Subnetworks representing the FL and SNF states were obtained by integrating transcriptome and proteome data [11,14] into the global PPI network of *B. diazoefficiens* USDA110. The subnetwork of the FL state contains 3650 proteins and 31,541 PPIs, while that of the SNF state contains 1777 proteins and 10,473 PPIs. In order to compare the topological differences between these two networks, five local metrics were calculated for each node. Interestingly, although the proportion of common proteins in the FL and SNF networks exceeds 80% of the SNF network, all the topological metrics are significantly different. The results (Table 1) show that the median of nodes' degrees is larger for the FL network due to its larger size, and that all the other metrics suggest the SNF network is more compact than the FL network, similar to the situation of the metabolic network of *B. diazoefficiens* USDA110 [76].

Table 1. Comparison of local topological properties between the PPI networks of the FL and SNF states (*p*-values were obtained by two-sided Wilcoxon rank sum test).

	Degree	Closeness Centrality	Betweenness Centrality	Average Shortest Path Length	Clustering Coefficient
Median for the FL network	9	0.322	4.35×10^{-5}	3.10	0.206
Median for the SNF network	7	0.328	7.54×10^{-5}	3.02	0.211
p-value	$<2.2 \times 10^{-16}$	2.87×10^{-4}	3.69×10^{-5}	6.01×10^{-6}	1.81×10^{-2}

Previous studies have shown that the two components of interacting protein pairs in the PPI network should have similar transcription levels [77], which may reflect the trade-off between the pairing efficiency and synthesis cost. Therefore, the normalized transcription level difference between the two components of each protein pair in the PPI network was calculated and analyzed based on the transcriptome data of *B. diazoefficiens* USDA110 [11,14,56]. First, the distribution of normalized transcription difference of the protein pairs in FL and SNF networks (real PPIs, the "PPIs" group) was compared with that of all pairwise combinations of proteins in the corresponding network (all possible protein pairs, the "Control" group), respectively. As shown in Figure 6A, the distribution of the "PPIs" group is lower than that of "Control" group in both the FL and SNF states (FL: *p*-value < 2.2×10^{-16}, median: 0.40 vs. 0.42, *n*: 31,541 vs. 13,322,500; SNF: *p*-value = 3.42×10^{-4}, median: 0.37 vs. 0.38, *n*: 10,473 vs. 3,157,729; compared by one-sided Wilcoxon rank sum test, $\alpha = 0.01$), which means that the transcription levels of the interacting proteins are more consistent. Furthermore, similar comparisons were performed in each COG group in the FL and SNF networks. The results showed that for the protein pairs belonging to the COG groups of C (energy production and conversion) and M (cell wall/membrane/envelope biogenesis), the normalized transcription differences of the FL state were significantly less than those of the SNF state, while for the COG groups of G (carbohydrate transport and metabolism), J (translation, ribosomal structure and biogenesis) and Q (secondary metabolites biosynthesis, transport and catabolism), the trend was the opposite (Figure 6B, Table S5).
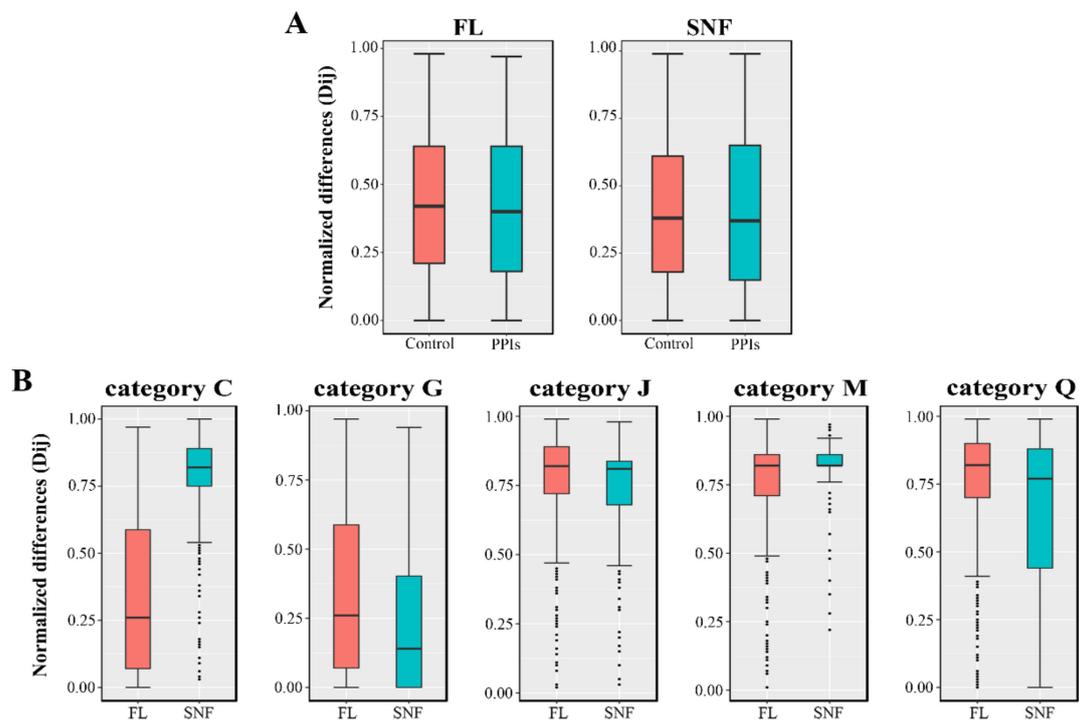
Figure 6. Comparison of normalized transcription level differences of interacting protein pairs. (**A**) Transcription differences in the Free-Living (FL) and Symbiotic Nitrogen Fixation (SNF) states. "Controls", all pairwise combinations for the proteins in the reconstructed PPI network; "PPIs", all the interacting protein pairs in the reconstructed PPI network. (**B**) Transcription differences in COG categories for the FL and SNF states. "Category C", the interacting protein pairs belonging to COG category C (energy production and conversion); and so on.

3.6. Analysis of the Core-Sub-PPI-Network for Symbiotic Nitrogen Fixation

Symbiotic nitrogen fixation is a complex physiological process involving many kinds of perfectly cooperated molecular machines. Functional genes related to symbiotic nitrogen fixation such as *nod*, *nif* and *fix*, and their regulation, have been studied intensively in the field of nitrogen-fixing bacteria, but their relationships with other physiological activities and relevant coordination are not very clear. Based on the reconstructed PPI network, functional modules of the core-sub-network related to SNF were identified and their cooperating relationships were analyzed.

First, a dataset of 128 SNF-associated proteins in *B. diazoefficiens* USDA110 was determined by genome annotation and literature mining (Table S6). Based on these SNF-associated proteins, corresponding PPIs were filtered out and the SNF core-sub-network (SCSNW) was determined, which contains 195 proteins and 441 PPIs (Supplementary Materials). Using a network clustering tool based on the Markov chain clustering (MCL) algorithm in the clusterMaker plug-in for Cytoscape, called MCL Cluster [78,79], a series of network modules of the SCSNW were obtained. As shown in Figure S2, most of the SNF associated-proteins are hubs of modules, which validates that the modules identified by MCL clustering are functionally meaningful. Then, nine key modules were selected and their functions were deduced by gene annotations from different databases such as KEGG and UniProtKB, as well as by literature mining. As shown in Table 2, all of the modules are associated with different aspects of the phenotype of SNF bacteroid. Furthermore, the whole nitrogen fixing system also depends on the coordination of the modules which can be achieved through the PPIs between their components. In addition to these PPIs, functional modules can also be connected through other hubs, which can be named 'Tie of Modules' (TOM). Hubs with specific properties (Figure S3) were selected as TOMs. Through the 11 TOMs, more diversified coordination could be established between functional mod-

ules (Figure 7A). Then the functions of PPIs binding TOMs and functional modules were deduced by genome annotation and literature mining (Table S7).

Table 2. Deduced functions of the network modules identified by MCL clustering.

| Modules	No. of Nodes	No. of PPIs	No. of SNF-Associated Proteins	Enriched COG Categories	Deduced Function
m1	24	49	2	O, K, E, F	Synthesis of symbiotic nitrogen fixation associated protein.
m2	18	40	4	T	Oxygen concentration response and transcriptional regulation of downstream N_2-fixing genes.
m3	12	29	3	E, C	Serine metabolism, transformation of amino-containing metabolites.
m4	12	28	3	T	Upstream regulation of nitrogenase and related transcription factors.
m5	11	33	7	E, C	N_2-fixing reaction, NH^{4+} transport and related regulation.
m6	11	32	4	T	DNA damage repair, oxygen stress defense
m7	7	7	1	E	Transformation of amino-containing metabolites.
m8	7	11	4	C	Energy production, electron transfer.
m9	6	8	2	C	Reduction of oxidative nitrogenase.
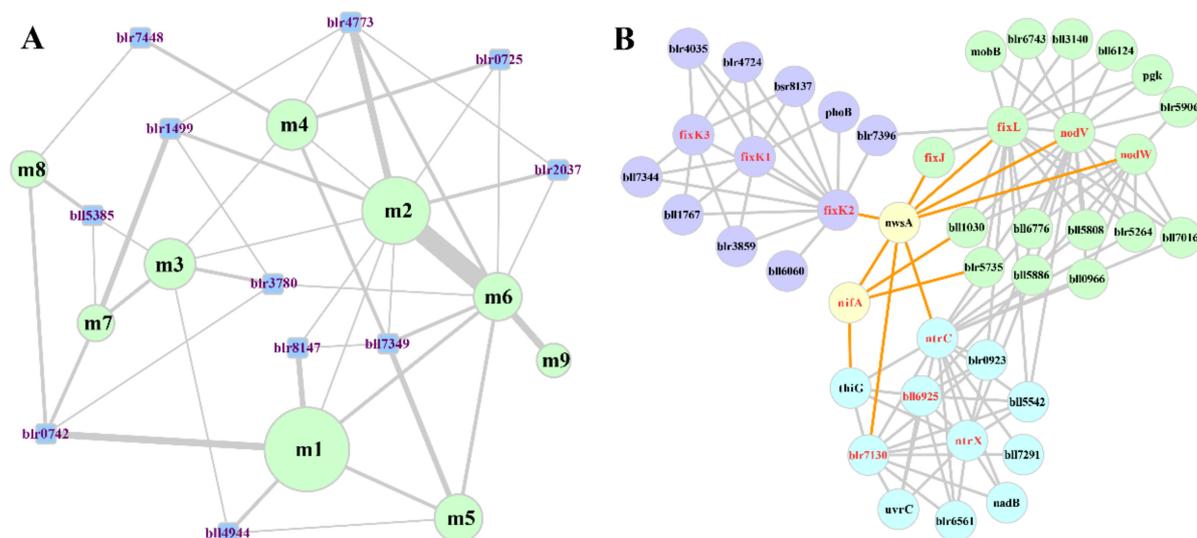
Figure 7. Subnetworks related to key modules (MCL clusters) and TOMs in the SNF core-sub-network. (**A**) Sub-network related to nine key modules (light-green nodes) and 11 TOMs (light-blue nodes). The diameter of a key module node corresponds to the number of proteins in this module. The width of an edge between two nodes corresponds to the number of PPIs. (**B**) Subnetwork related to NwsA (including its interacting modules and TOMs). Light-green, blue and purple nodes represent the m2, m4 and m6 modules, respectively. Nodes with red labels represent the SNF-associated proteins.

As shown in Figure 7B, we found protein NwsA, a two-component hybrid sensor and regulator coded by gene *blr4773*, connects three functional modules (m2, m4, m6) and interacts with several key hubs of regulatory cascades in the SNF bacteroid. As to these regulators, NodV and NodW regulate the expression of other *nod* genes (such as *nod*YABC) that initiate nodulation [80,81], the FixLJ-FixK2-FixK1 cascade activates genes essential for microoxic respiration in symbiosis (such as *fix*NOPQ and *fix*GHIS) and further regulatory genes for nitrogen fixation (*rpo*N1, *nnr*R, and *fix*K1) [82–84], NifA controls expression of nitrogen fixation genes (such as *nif*HDK) [83,85] and NtrC partly controls the regulation of genes related to nitrogen metabolism (such as *gln*II) [86,87]. Studies have

reported that cross talk between the FixLJ-FixK2-FixK1 cascade and RegSR-NifA cascade might allow switching between different expression patterns of genes essential for N_2 fixation [83,88]. Therefore, NwsA might play a critical role in the cross talk, and probably performs an important coordination function between nodulation, nitrogen fixation and nitrogen assimilation in the symbiotic nitrogen fixation system. This waits for experimental validation. Just like NwsA, other TOMs might also be important for symbiotic nitrogen fixation through interactions with related proteins, although further information needs to be acquired.

4. Conclusions

In this work, the protein interactome (60,839 PPIs among 5638 proteins) of *B. diazoefficiens* USDA110 was computationally reconstructed by combining Interolog and domain-based methods, and its reliability was validated from four perspectives. The reconstructed interactome of *B. diazoefficiens* USDA110 was compared with those of other rhizobia (*M. loti* and *S. meliloti*) and it was inferred that the slow growth of *B. diazoefficiens* USDA110 may require more protein modification and repair via protein-protein interaction. By network comparison, 36 conserved functional modules were identified, and the functions of related proteins were annotated. By integrating the reconstructed interactome with transcriptome and proteome data, the subnetworks representing FL and SNF states of *B. diazoefficiens* USDA110 were derived. Based on the SNF subnetwork and the SNF-associated proteins, nine novel functional modules and eleven protein hubs (named TOMs which connect functional modules) were identified and analyzed to further our understanding of the molecular mechanism of symbiotic nitrogen fixation.

Abbreviations

PPI, protein-protein interaction; SNF, symbiotic nitrogen-fixing; FL, free-living; DDI, domain-domain interaction; GO, gene ontology; PCC, Pearson correlation coefficient; COG, clusters of orthologous groups of proteins; GEO, gene expression omnibus; MDS, minimum dominating set; 3DID, three-dimensional interacting domains; MCL, Markov chain clustering; TOM, Tie of Modules; SCSNW, the SNF core-sub-network.

References

1. Galloway, J.N. An Earth-system perspective of the global nitrogen cycle. *Nature* **2008**, *451*, 293–296.
2. Delamuta, J.R.M.; Ribeiro, R.A.; Ormeño-Orrillo, E.; Melo, I.S.; Martínez-Romero, E.; Hungria, M. Polyphasic evidence supporting the reclassification of *Bradyrhizobium japonicum* group Ia strains as *Bradyrhizobium diazoefficiens* sp. nov. *Int. J. Syst. Evol. Microbiol.* **2013**, *63*, 3342–3351. [CrossRef]
3. Hahn, M.; Hennecke, H. Mapping of a *Bradyrhizobium japonicum* DNA Region region carrying genes for symbiosis and an asymmetric accumulation of reiterated sequences. *Appl. Environ. Microbiol.* **1987**, *53*, 2247. [CrossRef] [PubMed]
4. Gillette, W.K.; Elkan, G.H. *Bradyrhizobium* (Arachis) sp. strain NC92 contains two nodD genes involved in the repression of nodA and a nolA gene required for the efficient nodulation of host plants. *J. Bacteriol.* **1996**, *178*, 2757–2766. [CrossRef]
5. Israel, D.W. Cultivar and *Rhizobium* strain effects on nitrogen fixation and remobilization by soybeans. *Agron. J.* **1981**, *73*, 509–516. [CrossRef]
6. Hunter, W.J.; Kuykendall, L.D. Enhanced nodulation and nitrogen fixation by a revertant of a nodulation-defective *Bradyrhizobium japonicum* tryptophan auxotroph. *Appl. Environ. Microbiol.* **1990**, *56*, 2399. [CrossRef]
7. Kuykendall, L.D.; Hashem, F.M.; Hunter, W.J. Enhanced competitiveness of a *Bradyrhizobium japonicum* mutant strain improved for nodulation and nitrogen fixation. *Plant Soil* **1996**, *186*, 121–125. [CrossRef]
8. Kosslak, R.M.; Bookland, R.; Barkei, J.; Paaren, H.E.; Appelbaum, E.R. Induction of *Bradyrhizobium japonicum* common nod genes by isoflavones isolated from *Glycine max*. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 7428–7432. [CrossRef]
9. Subramanian, S.; Stacey, G.; Yu, O. Endogenous isoflavones are essential for the establishment of symbiosis between soybean and *Bradyrhizobium japonicum*. *Plant J.* **2006**, *48*, 261–273. [CrossRef] [PubMed]
10. Kaneko, T.; Nakamura, Y.; Sato, S.; Minamisawa, K.; Uchiumi, T.; Sasamoto, S.; Watanabe, A.; Idesawa, K.; Iriguchi, M.; Kawashima, K.; et al. Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110 (Supplement). *DNA Res.* **2002**, *9*, 225–256. [CrossRef] [PubMed]
11. Pessi, G.; Ahrens, C.H.; Rehrauer, H.; Lindemann, A.; Hauser, F.; Fischer, H.M.; Hennecke, H. Genome-wide transcript analysis of *Bradyrhizobium japonicum* bacteroids in soybean root nodules. *Mol. Plant Microbe Interact.* **2007**, *20*, 1353–1363. [CrossRef]
12. Sarma, A.D.; Emerich, D.W. Global protein expression pattern of *Bradyrhizobium japonicum* bacteroids: A prelude to functional proteomics. *Proteomics* **2005**, *5*, 4170. [CrossRef] [PubMed]
13. Sarma, A.D.; Emerich, D.W. A comparative proteomic evaluation of culture grown vs nodule isolated *Bradyrhizobium japonicum*. *Proteomics* **2006**, *6*, 3008–3028. [CrossRef]
14. Delmotte, N.; Ahrens, C.H.; Knief, C.; Qeli, E.; Koch, M.; Fischer, H.M.; Vorholt, J.A.; 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. [CrossRef] [PubMed]
15. Jelena, Č.; Hahn, J.; Imakaev, M.; Omasits, U.; Förstner, K.U.; Ljubimov, N.; Goebel, M.; Pessi, G.; Fischer, H.M.; Ahrens, C.H.; 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 Genom.* **2016**, *17*, 302.
16. Lardi, M.; Murset, V.; Fischer, H.M.; Mesa, S.; Ahrens, C.H.; Zamboni, N.; Pessi, G. Metabolomic profiling of *Bradyrhizobium diazoefficiens*-induced root nodules reveals both host plant-specific and developmental signatures. *Int. J. Mol. Sci.* **2016**, *17*, 815. [CrossRef]
17. Zhang, L.; Liu, J.; Gu, H.; Du, Y.; Zuo, J.; Zhang, Z.; Zhang, M.; Li, P.; Dunwell, J.M.; Cao, Y.; et al. *Bradyrhizobium diazoefficiens* USDA 110-Glycine max interactome provides candidate proteins associated with symbiosis. *J. Proteome Res.* **2018**, *17*, 3061–3074. [CrossRef] [PubMed]
18. Gavin, A.C.; Supertifurga, G. Protein complexes and proteome organization from yeast to man. *Curr. Opin. Chem. Biol.* **2003**, *7*, 21–27. [CrossRef]
19. Bauer, A.; Kuster, B. Affinity purification-mass spectrometry. Powerful tools for the characterization of protein complexes. *Eur. J. Biochem.* **2003**, *270*, 570–578. [CrossRef] [PubMed]
20. Mosca, R.; Pons, T.; Céol, A.; Valencia, A.; Aloy, P. Towards a detailed atlas of protein-protein interactions. *Curr. Opin. Struct. Biol.* **2013**, *23*, 929–940. [CrossRef]
21. Taylor, I.W.; Linding, R.; Farley, D.W.; Liu, Y.; Pesquita, C.; Faria, D.; Bull, S.; Pawson, T.; Morris, Q.; Wrana, J.L. Dynamic modularity in protein interaction networks predicts breast cancer outcome. *Nat. Biotechnol.* **2009**, *27*, 199–204. [CrossRef] [PubMed]
22. Han, J.D.; Bertin, N.; Hao, T.; Goldberg, D.S.; Berriz, G.F.; Zhang, L.V.; Dupuy, D.; Walhout, A.J.; Cusick, M.E.; Roth, F.P.; et al. Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature* **2004**, *430*, 88. [CrossRef] [PubMed]
23. Wuchty, S. Controllability in protein interaction networks. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 7156. [CrossRef] [PubMed]
24. Schwikowski, B.; Uetz, P.; Fields, S. A network of protein-protein interactions in yeast. *Nat. Biotechnol.* **2000**, *18*, 1257. [CrossRef] [PubMed]
25. Mostafavi, S.; Ray, D.; Warde-Farley, D.; Grouios, C.; Morris, Q. GeneMANIA: A real-time multiple association network integration algorithm for predicting gene function. *Genome Biol.* **2008**, *9* (Suppl. 1), S4. [CrossRef]
26. Giot, L.; Bader, J.S.; Brouwer, C.; Chaudhuri, A.; Kuang, B.; Li, Y.; Hao, Y.L.; Ooi, C.E.; Godwin, B.; Vitols, E.; et al. A protein interaction map of *Drosophila melanogaster*. *Science* **2003**, *302*, 1727–1736. [CrossRef]

27. Uetz, P.; Giot, L.; Cagney, G.; Mansfield, T.A.; Judson, R.S.; Knight, J.R.; Lockshon, D.; Narayan, V.; Srinivasan, M.; Pochart, P.; et al. A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* **2000**, *403*, 623. [CrossRef] [PubMed]

28. Hao, T.; Peng, W.; Wang, Q.; Wang, B.; Sun, J. Reconstruction and application of protein–protein interaction network. *Int. J. Mol. Sci.* **2016**, *17*, 907. [CrossRef]

29. Von Mering, C.; Krause, R.; Snel, B.; Cornell, M.; Oliver, S.G.; Fields, S.; Bork, P. Comparative assessment of large-scale data sets of protein-protein interactions. *Nature* **2002**, *417*, 399. [CrossRef]

30. Stumpf, M.P.H.; Thorne, T.; de Silva, E.; Stewart, R.; An, H.J.; Lappe, M.; Wiuf, C. Estimating the size of the human interactome. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 6959–6964. [CrossRef]

31. Yu, J.; Guo, M.; Needham, C.J.; Huang, Y.; Cai, L.; Westhead, D.R. Simple sequence-based kernels do not predict protein–protein interactions. *Bioinformatics* **2010**, *26*, 2610–2614. [CrossRef] [PubMed]

32. Zhang, X.; Xiong, J.; Song, J.; Chang, S. Prediction of human protein–protein interaction by a domain-based approach. *J. Theor. Biol.* **2016**, *396*, 144–153. [CrossRef] [PubMed]

33. Kotlyar, M.; Aem, R.; Jurisica, I. Prediction of protein-protein interactions. *Curr. Protoc. Bioinform.* **2017**, *60*, 8.2.1–8.2.14. [CrossRef] [PubMed]

34. Walhout, A.J.; Walhout, A.J.; Sordella, R.; Lu, X.; Hartley, J.L.; Temple, G.F.; Brasch, M.A.; Thierry-Mieg, N.; Vidal, M. Protein interaction mapping in *C. elegans* using proteins involved in vulval development. *Science* **2000**, *287*, 116–122. [CrossRef]

35. Matthews, L.R.; Vaglio, P.; Reboul, J.; Ge, H.; Davis, B.P.; 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. [CrossRef] [PubMed]

36. Parrish, J.R.; Yu, J.; Liu, G.; Hines, J.A.; Chan, J.E.; Mangiola, B.A.; Zhang, H.; Pacifico, S.; Fotouhi, F.; DiRita, V.J.; et al. A proteome-wide protein interaction map for *Campylobacter jejuni*. *Genome Biol.* **2007**, *8*, R130. [CrossRef]

37. Rajagopala, S.V.; Sikorski, P.; Kumar, A.; Mosca, R.; Vlasblom, J.; Arnold, R.; Franca-Koh, J.; Pakala, S.B.; Phanse, S.; Ceol, A.; et al. The binary protein-protein interaction landscape of *Escherichia coli*. *Nat. Biotechnol.* **2014**, *34*, 285–290. [CrossRef]

38. Häuser, R.; Ceol, A.; Rajagopala, S.V.; Mosca, R.; Siszler, G.; Wermke, N.; Sikorski, P.; Schwarz, F.; Schick, M.; Wuchty, S.; et al. A second-generation protein–protein interaction network of *Helicobacter pylori*. *Mol. Cell. Proteom.* **2014**, *13*, 1318. [CrossRef]

39. Shimoda, Y.; Shinp, S.; Kohara, M.; Nakamura, Y.; Tabata, S.; Sato, S. A large scale analysis of protein-protein interactions in the nitrogen-fixing bacterium *Mesorhizobium loti*. *DNA Res.* **2008**, *15*, 13–23. [CrossRef] [PubMed]

40. Sato, S.; Shimoda, Y.; Muraki, A.; Kohara, M.; Nakamura, Y.; Tabata, S. A large-scale protein protein interaction analysis in *Synechocystis* sp. PCC6803. *DNA Res.* **2007**, *14*, 207. [CrossRef]

41. Titz, B.; Rajagopala, S.V.; Goll, J.; Häuser, R.; McKevitt, M.T.; Palzkill, T.; Uetz, P. The binary protein interactome of *Treponema pallidum*-the syphilis spirochete. *PLoS ONE* **2008**, *3*, e2292. [CrossRef] [PubMed]

42. Stark, C.; Breitkreutz, B.J.; Reguly, T.; Boucher, L.; Breitkreutz, A.; Tyers, M. BioGRID: A general repository for interaction datasets. *Nucleic Acids Res.* **2006**, *34*, D535–D539. [CrossRef] [PubMed]

43. Xenarios, I.; Rice, D.W.; Salwinski, L.; Baron, M.K.; Marcotte, E.M.; Eisenberg, D. DIP: The database of interacting proteins. *Nucleic Acids Res.* **2000**, *28*, 289–291. [CrossRef] [PubMed]

44. Orchard, S.; Ammari, M.; Aranda, B.; Breuza, L.; Briganti, L.; Broackes-Carter, F.; Campbell, N.H.; Chavali, G.; Chen, C.; del-Toro, N.; et al. The MIntAct project-IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res.* **2014**, *42*, D358–D363. [CrossRef]

45. Peri, S.; Navarro, J.D.; Amanchy, R.; Kristiansen, T.Z.; Jonnalagadda, C.K.; Surendranath, V.; Niranjan, V.; Muthusamy, B.; Gandhi, T.K.; Gronborg, M.; et al. Development of human protein reference database as an initial platform for approaching systems biology in humans. *Genome Res.* **2003**, *13*, 2363–2371. [CrossRef] [PubMed]

46. Remm, M.; Storm, C.E.V.; Sonnhammer, E.L.L. Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *J. Mol. Biol.* **2001**, *314*, 1041–1052. [CrossRef]

47. Mosca, R.; Ceol, A.; Stein, A.; Olivella, R.; Aloy, P. 3did: A catalog of domain-based interactions of known three-dimensional structure. *Nucleic Acids Res.* **2014**, *42*, D374–D379. [CrossRef] [PubMed]

48. Finn, R.D.; Bateman, A.; Clements, J.; Coggill, P.; Eberhardt, R.Y.; Eddy, S.R.; Heger, A.; Hetherington, K.; Holm, L.; Mistry, J.; et al. Pfam: The protein families database. *Nucleic Acids Res.* **2014**, *42*, D222–D230. [CrossRef]

49. Finn, R.D.; Clements, J.; Eddy, S.R. HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Res.* **2011**, *39*, W29–W37. [CrossRef]

50. Planas-Iglesias, J.; Marin-Lopez, M.A.; Bonet, J.; Garcia-Garcia, J.; Oliva, B. iLoops: A protein-protein interaction prediction server based on structural features. *Bioinformatics* **2013**, *29*, 2360–2362. [CrossRef]

51. Planas-Iglesias, J.; Bonet, J.; García-García, J.; Marín-López, M.A.; Feliu, E.; Oliva, B. Understanding protein-protein interactions using local structural features. *J. Mol. Biol.* **2013**, *425*, 1210–1224. [CrossRef] [PubMed]

52. Yu, N.Y.; Wagner, J.R.; Laird, M.R.; Melli, G.; Rey, S.; Lo, R.; Dao, P.; Sahinalp, S.C.; Ester, M.; Foster, L.J.; et al. PSORTb 3.0: Improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* **2010**, *26*, 1608–1615. [CrossRef]

53. Wang, J.Z.; Du, Z.; Payattakool, R.; Yu, P.S.; Chen, C.F. A new method to measure the semantic similarity of GO terms. *Bioinformatics* **2007**, *23*, 1274–1281. [CrossRef] [PubMed]

54. Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T.; et al. Gene Ontology: Tool for the unification of biology. *Nat. Genet.* **2000**, 25, 25–29. [CrossRef]

55. Edgar, R.; Domrachev, M.; Lash, A.E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **2002**, 30, 207–210. [CrossRef] [PubMed]

56. Jansen, R.; Greenbaum, D.; Gerstein, M. Relating whole-genome expression data with protein-protein interactions. *Genome Res.* **2002**, 12, 37. [CrossRef]

57. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, 13, 2498–2504. [CrossRef]

58. Tatusov, R.L.; Koonin, E.V.; Lipman, D.J. A genomic perspective on protein families. *Science* **1997**, 278, 631–637. [CrossRef]

59. Lehner, B.; Fraser, A.G. A first-draft human protein-interaction map. *Genome Biol.* **2004**, 5, R63. [CrossRef]

60. Ge, H.; Liu, Z.; Church, G.M.; Vidal, M. Correlation between transcriptome and interactome mapping data from *Saccharomyces cerevisiae*. *Nat. Genet.* **2001**, 29, 482. [CrossRef]

61. Assenov, Y.; Ramírez, F.; Schelhorn, S.E.; Lengauer, T.; Albrecht, M. Computing topological parameters of biological networks. *Bioinformatics* **2008**, 24, 282–284. [CrossRef] [PubMed]

62. Wolf, Y.I.; Karev GKoonin, E.V. Scale-free networks in biology: New insights into the fundamentals of evolution? *Bioessays* **2002**, 24, 105–109. [CrossRef]

63. Mirmiran, A.; Desveaux, D.; Subramaniam, R. Building a protein-interaction network to study *Fusarium graminearum* pathogenesis. *Can. J. Plant Pathol.* **2018**, 40, 172–178. [CrossRef]

64. Pang, E.; Tan, T.; Lin, K. Promiscuous domains: Facilitating stability of the yeast protein-protein interaction network. *Mol. Biosyst.* **2012**, 8, 766–771. [CrossRef] [PubMed]

65. Nacher, J.C.; Akutsu, T. Dominating scale-free networks with variable scaling exponent: Heterogeneous networks are not difficult to control. *N. J. Phys.* **2012**, 14, 073005. [CrossRef]

66. Peregrínalvarez, J.M.; Xiong, X.; Chong, S.; Parkinson, J. The modular organization of protein interactions in *Escherichia coli. PLoS Comput. Biol.* **2009**, 5, e1000523.

67. Rodriguez-Llorente, I.; Caviedes, M.A.; Dary, M.; Palomares, A.J.; Cánovas, F.M.; Peregrín-Alvarez, J.M. The Symbiosis Interactome: A computational approach reveals novel components, functional interactions and modules in *Sinorhizobium meliloti. BMC Syst. Biol.* **2009**, 3, 63. [CrossRef]

68. Dunn, M.F. Key roles of microsymbiont amino acid metabolism in rhizobia-legume interactions. *Crit. Rev. Microbiol.* **2015**, 41, 411–451. [CrossRef]

69. Sharan, R.; Suthram, S.; Kelley, R.M.; Kuhn, T.; McCuine, S.; Uetz, P.; Sittler, T.; Karp, R.M.; Ideker, T. Conserved patterns of protein interaction in multiple species. *Proc. Natl. Acad. Sci. USA* **2005**, 102, 1974–1979. [CrossRef]

70. Davis, D.; Yaveroğlu, Ö.N.; Malod-Dognin, N.; Stojmirovic, A.; Pržulj, N. Topology-function conservation in protein-protein interaction networks. *Bioinformatics* **2015**, 31, 1632–1639. [CrossRef]

71. Micale, G.; Pulvirenti, A.; Giugno, R.; Ferro, A. GASOLINE: A greedy and stochastic algorithm for optimal local multiple alignment of interaction networks. *PLoS ONE* **2014**, 9, e98750. [CrossRef] [PubMed]

72. Quelas, J.I.; Althabegoiti, M.J.; Jimenez-Sanchez, C.; Melgarejo, A.A.; Marconi, V.I.; Mongiardini, E.J.; Trejo, S.A.; Mengucci, F.; Ortega-Calvo, J.J.; Lodeiro, A.R. Swimming performance of *Bradyrhizobium diazoefficiens* is an emergent property of its two flagellar systems. *Sci. Rep.* **2016**, 6, 23841. [CrossRef] [PubMed]

73. Kumar, K.; Mellaherrera, R.A.; Golden, J.W. Cyanobacterial heterocysts. *Cold Spring Harb. Perspect. Biol.* **2010**, 2, a000315. [CrossRef] [PubMed]

74. Bhargava, S.; Chouhan, S. Diazotrophic specific cytochrome c oxidase required to overcome light stress in the cyanobacterium *Nostoc muscorum. World J. Microbiol. Biotechnol.* **2016**, 32, 2. [CrossRef]

75. Xu, T.; Li, H.; Lu, S.; Zhang, H.; Ge, J. Research progress of mitochondrial electron transport chain and its biological significance. *Fudan Univ. J. Med. Sci.* **2015**, 42, 250–261.

76. Yang, Y.; Hu, X.P.; Ma, B.G. Construction and simulation of the *Bradyrhizobium diazoefficiens* USDA110 metabolic network: A comparison between free-living and symbiotic states. *Mol. Biosyst.* **2017**, 13, 607–620. [CrossRef]

77. Bhardwaj, N.; Lu, H. Correlation between gene expression profiles and protein–protein interactions within and across genomes. *Bioinformatics* **2005**, 21, 2730–2738. [CrossRef]

78. Morris, J.H.; Apeltsin, L.; Newman, A.M.; Baumbach, J.; Wittkop, T.; Su, G.; Bader, G.D.; Ferrin, T.E. ClusterMaker: A multi-algorithm clustering plugin for Cytoscape. *BMC Bioinform.* **2011**, 12, 436. [CrossRef]

79. Enright, A.J.; Van, D.S.; Ouzounis, C.A. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* **2002**, 30, 1575–1584. [CrossRef]

80. Grob, P.; Michel, P.; Hennecke, H.; Göttfert, M. A novel response-regulator is able to suppress the nodulation defect of a *Bradyrhizobium japonicum* nodW mutant. *Mol. Gen. Genet.* **1993**, 241, 531–541. [CrossRef]

81. Loh, J.; Stacey, G. Nodulation gene regulation in *Bradyrhizobium japonicum*: A unique integration of global regulatory circuits. *Appl. Environ. Microbiol.* **2003**, 69, 10–17. [CrossRef] [PubMed]

82. Nellen-anthamatten, D.; Rossi, P.; Preisig, O.; Kullik, I.; Babst, M.; Fischer, H.M.; Hennecke, H. *Bradyrhizobium japonicum* FixK2, a crucial distributor in the FixLJ-dependent regulatory cascade for control of genes inducible by low oxygen levels. *J. Bacteriol.* **1998**, 180, 5251–5255. [CrossRef] [PubMed]

83. Dixon, R.; Kahn, D. Genetic regulation of biological nitrogen fixation. *Nat. Rev. Microbiol.* **2004**, *2*, 621–631. [CrossRef]

84. Mesa, S.; Hauser, F.; Friberg, M.; Malaguti, E.; Fischer, H.M.; Hennecke, H. Comprehensive assessment of the regulons controlled by the FixLJ-FixK2-FixK1 cascade in *Bradyrhizobium japonicum*. *J. Bacteriol.* **2008**, *190*, 6568–6579. [CrossRef]

85. Lindemann, A.; Moser, A.; Pessi, G.; Hauser, F.; Friberg, M.; Hennecke, H.; Fischer, H.M. New target genes controlled by the *Bradyrhizobium japonicum* two-component regulatory system RegSR. *J. Bacteriol.* **2007**, *189*, 8928. [CrossRef] [PubMed]

86. Martin, G.B.; Chapman, K.A.; Chelm, B.K. Role of the *Bradyrhizobium japonicum ntr*C gene product in differential regulation of the glutamine synthetase II gene (*gln*II). *J. Bacteriol.* **1988**, *170*, 5452–5459. [CrossRef]

87. Franck, W.L.; Qiu, J.; Lee, H.I.; Chang, W.S.; Stacey, G. DNA microarray-based identification of genes regulated by NtrC in *Bradyrhizobium japonicum*. *Appl. Environ. Microbiol.* **2015**, *81*, 5299–5308. [CrossRef] [PubMed]

88. Terpolilli, J.J.; Hood, G.A.; Poole, P.S. What determines the efficiency of N$_2$-fixing *Rhizobium*-legume symbioses? *Adv. Microb. Physiol.* **2012**, *60*, 325.