A Large Scale Analysis of Protein−Protein Interactions in the Nitrogen-fixing Bacterium *Mesorhizobium loti*

Yoshikazu Shimoda, Sayaka Shinpo, Mitsuyo Kohara, Yasukazu Nakamura, Satoshi Tabata, and Shusei Sato*

Kazusa DNA Research Institute, 2-6-7 Kazusa-kamatari, Kisarazu, Chiba 292-0818, Japan

Abstract

   Global viewing of protein−protein interactions (PPIs) is a useful way to assign biological roles to large numbers of proteins predicted by complete genome sequence. Here, we systematically analyzed PPIs in the nitrogen-fixing soil bacterium *Mesorhizobium loti* using a modified high-throughput yeast two-hybrid system. The aims of this study are primarily on the providing functional clues to *M. loti* proteins that are relevant to symbiotic nitrogen fixation and conserved in other rhizobium species, especially proteins with regulatory functions and unannotated proteins. By the screening of 1542 genes as bait, 3121 independent interactions involving 1804 proteins (24% of the total protein coding genes) were identified and each interaction was evaluated using an interaction generality (IG) measure and the general features of the interacting partners. Most PPIs detected in this study are novel interactions revealing potential functional relationships between genes for symbiotic nitrogen fixation and signal transduction. Furthermore, we have predicted the putative functions of unannotated proteins through their interactions with known proteins. The results described here represent new insight into protein network of *M. loti* and provide useful experimental clues to elucidate the biological function of rhizobial genes that can not be assigned directly from their genomic sequence.

Key words: *Mesorhizobium loti*; Protein−protein interaction; Yeast two-hybrid system

1.   Introduction

   Bacteria belonging to the family Rhizobiaceae (rhizobia) are biologically unique organisms that exhibit two distinct lifestyles in nature, as free-living soil bacteria or as symbionts of some leguminous plants. In nitrogen-starved soil, rhizobia can colonize the roots of compatible legumes and elicit the formation of specialized organ, the root nodules. Inside the nodules, rhizobia differentiate into non-dividing form of bacteria (bacteroid) and reduce atmospheric dinitrogen (N_2) into biologically usable ammonia (NH_3). Nitrogen fixed by the symbiotic rhizobia is then assimilated by the host plant, enabling the plant to grow in nitrogen-depleted environments.

   Because of the agronomic importance of rhizobia, complete genome sequencing has been accomplished in seven rhizobia[1−6] and is now in progress on several additional species. The availability of complete genome sequences allows us to know the global features of the organism 'rhizobium' at the molecular level. Simultaneously, the genome sequences also provide genetic platforms for comparative study of genome structure within rhizobia, or between rhizobia and other plant-associated bacteria. These comparative analyses have revealed the existence of a genomic region enriched for genes involved in symbiotic nitrogen fixation (symbiosis island), as well as various specialized gene sets for signal transduction and membrane transport that allow rhizobia to

---

adapt to the environment of both soil and intracellular of host plant.[1,5,7]

Although a genome sequence is very useful, the genome sequence alone is not sufficient to reveal the specific molecular functions of all the genes. To date, genetic analyses, such as targeted gene disruption or transposon random mutagenesis, have revealed various rhizobial genes essential for nodule formation and nitrogen fixation.[6,8] Additionally, with the completion of several rhizobial genome sequences, comprehensive transcriptome or proteome analysis has been carried out in order to understand physiological states of rhizobia under a variety of conditions such as symbiosis with host legume or nutrient-depleted conditions.[9–14] These analyses have allowed us to know expressional dynamics throughout the rhizobial genome and to discover numerous novel genes and proteins not previously known to be involved in these conditions.

Although a large amount of data have been accumulated through these functional analyses, many genes predicted on the rhizobial genome remain functionally unannotated. That is primarily because that the genome sequence and expression profiles give an indirect and fragmented picture of the biological function of genes, and more detailed information on gene function would be obtained by biochemical properties of the gene products (proteins) or by their interaction with other proteins of known function. Therefore, more interactive and systematic analysis that accelerates functional assignment for many rhizobial genes simultaneously is required.

As a first effort to examine the protein networks of rhizobia, we conducted a large-scale analysis of protein–protein interactions (PPIs) in *Mesorhizobium loti* using a yeast two-hybrid (YTH) system. YTH analysis is one of the well-established methods for mapping binary protein interactions and its usefulness is clearly illustrates by the fact that comprehensive studies of PPIs in several model organisms have provided several important biological and bioinformatics platforms for the study of protein networks in the organism.[15–18] These analyses have also successfully placed functionally uncharacterized proteins into their biological context.

In this study, for the selection of target gene groups for the YTH screening, we primarily used two available information, the genomic information and the results of a transcriptome analysis of *M. loti*. Symbiosis island contains many genes for symbiotic nitrogen fixation[1] and the clustered up-regulation of genes in symbiosis island under bacteroid state were observed by macroarray analysis of *M. loti*.[14] The macroarray analysis also revealed that several genes located outside of symbiosis island were also up-regulated under bacteroid or microaerobic condition. Although these features led us to hypothesize that genes located in the symbiotic island and those up-regulated under symbiosis are involved in certain physiological events of symbiotic nitrogen fixation, the biological significance of the majority of these genes has not yet been revealed. Identification of PPIs of these gene products will help elucidate their biological function and expand our knowledge of the mechanism of symbiotic nitrogen fixation. Therefore, we selected these genes for the initial target of our analysis. Complete genome sequences of several rhizobium species enables us to survey the conserved orthologous genes among rhizobia. Since conserved genes are presumed to execute common functions, the obtained PPI data from the *M. loti* YTH screening can be applied to other rhizobia. Among conserved genes, we primarily selected *M. loti* genes of unknown function and those with regulatory functions, in order to reveal the biological roles of the functionally unannotated proteins and signal transduction pathways that function commonly among rhizobium species.

2. Materials and methods

2.1. Construction of bait clones and prey library

The backbone of our YTH screening is the MATCHMAKER GAL-4 based YTH system. As for bait and prey vectors, we used pAS2-1 and pACT2 (Clontech, Mountain View, CA, USA), respectively, which were modified by introducing the Gateway recombination system (Invitrogen, Carlsbad, CA, USA).[19] A Gateway cassette containing the attR recombination site flanking a *ccd*B gene and a chloramphenicol-resistance gene were ligated into the multicloning site of pAS2-1 and pACT2, and the resultant vectors were designated as pAS-GW and pACT-GW, respectively.

For bait clones, target gene fragments were obtained by PCR amplification from cosmid clones or genomic DNA using gene specific primer pairs with CACC sequence added to the forward primer. To minimize potential misincorporation during PCR, the high-fidelity DNA polymerase, *Pfu* DNA polymerase (Stratagene, La Jolla, CA, USA), was used. Amplified fragments were cloned into the pENTR/D-TOPO vector (Invitrogen) to make entry clones. After confirmation of DNA sequence, the insert of the entry clone was transferred to pAS-GW using LR recombination reaction according to the manufacturer's instructions. The resultant plasmids harboring individual *M. loti* genes were transformed into yeast AH109 (*MATa*) (Clontech) and transformants were grown on SD/-Trp plates.

We constructed a prey library from a random genomic fragment library. Genomic DNA of *M. loti* was isolated by the procedure of Chen and Kuo[20]

with some modifications. Following the sonication of genomic DNA, two size ranges of DNA fragments (0.5−1.2 and 1.2−2.5 kb) were isolated from agarose gels and then cloned individually into *Dra*I/*Eco*RV site of the pENTR 1A vector (Invitrogen). A pool of plasmids containing each size range of DNA fragments were recovered from 5.3×10^6 (0.5−1.2 kb) and 3.8×10^6 (1.2−2.5 kb) independent *Escherichia coli* transformants (ElectroMax DH10B competent cells; Invitrogen). The inserts of the recovered plasmids were transferred to the pACT-GW vector via LR recombination, and the resultant plasmids were transformed into the yeast strain Y187 (*MATα*). A total of 6×10^6 yeast transformants were collected and pooled as the prey library.

2.2. Mating-mediated YTH screening

Prior to screening, the self-activity of each bait clone was confirmed by mating with the Y187 strain harboring an empty pACT2 vector and then plating on SD/-His/-Leu/-Trp/ medium supplemented with 2.5, 5, 10, or 50 mM 3-amino 1,2,4-triazole (3-AT). Each bait clone was mated with the prey library containing approximately 3×10^7 independent clones and plated on SD/-His/-Leu/-Trp/ agar medium supplemented with the optimal concentration of 3-AT. After 7 days of growth at 30°C, positive colonies were picked and transferred into 96-well culture plates and grown for an additional 3 days at 30°C. Some of each cultured positive clone was used for β-galactosidase assays and DNA sequencing of the prey clones insert, and the remainder was stored at −80°C. The collected positive clones were treated with Zymolyase solution [2.5 mg/ml of Zymolyase-100T; (Seikagaku America Inc., EastFalmouth, MA, USA), 1.2 M sorbitol, 0.1 M Na phosphate, pH 7.4] for 30 min at 37°C and then used as templates for amplification of prey clone inserts using the following primers: 5′-TACCACTACAATGGATGATG-3′ and 5′-GGGGTTTTTCAGTATCTACG-3′. Amplified fragments were sequenced using the same primers to obtain sequence tags. The resulting sequences were compared with the RhizoBase (http://www.kazusa.or.jp/rhizobase) to determine the genomic region which interacted with each bait protein.

2.3. Homolog search and calculation of interaction generality

Mesorhizobium loti genes whose homologs are conserved in three rhizobia were selected according to the criteria described by Kaneko et al.[2] The lower threshold of acceptability was set at 0.25 of the BLASTP bit score reported by self-comparison. Paralogous genes of *M. loti* were selected if the Smith−Waterman score was greater than 200.

The calculation of the interaction generality (IG) values of each interaction was conducted using the methods described in the previous report.[21]

3. Results and discussion

3.1. Design of a high-throughput YTH system and selection of M. loti genes for screening

To facilitate the efficiency and accuracy of directional in-frame cloning of *M. loti* genes, we used a Gateway-compatible vector system to construct bait clones (Fig. 1). Using this system, we successfully cloned 1542 full-length *M. loti* genes (21% of the total *M. loti* genes, Supplementary Table 1) into the Gateway entry vector. Since the insert of each entry clone can be easily transferred to other destination vectors via Gateway recombination, the constructed Gateway entry clones can also be used as material resources for many other analyses of *M. loti* genes.

In order to screen various potential PPIs efficiently, we constructed prey library from GAL4 activation domain-fused random genomic library and screened the library using a yeast-mating method. By using the random fragment library as prey clones, information on interaction pairs and interacting regions of the prey proteins can be obtained simultaneously. All selected positive clones were processed in 96-well plate format and data of a large number of sequence tags derived from positive clones were processed by the same semi-automated system developed by Sato et al.[22]

Using this high-throughput YTH system, we explored large-scale PPI analysis in the nitrogen-fixing soil bacteria, *M. loti*. As shown in Table 1, we selected *M. loti* genes for YTH screening on the basis of the following features: (i) genes whose expression is up-regulated under bacteroid or microaerobic condition and genes located in symbiosis island, and (ii) genes whose homologs are conserved in other rhizobium genomes.

Up-regulated genes were selected by referring to the macroarray analysis of *M. loti*.[14] On the basis of the supplemental information provided with the macroarray analysis, 93 and 72 genes were selected as highly up-regulated under bacteroid state and microaerobic condition, respectively (Supplementary Table 1). Genes located in symbiosis island were selected from genomic information of *M. loti*.[1] We selected 416 genes (excluding those genes with transposon-related functions and those previously selected as up-regulated under bacteroid or microaerobic condition) from the location of symbiosis island (Supplementary Table 1). Genes whose homologs are conserved in other rhizobium genomes were selected by comparing the genome sequences of
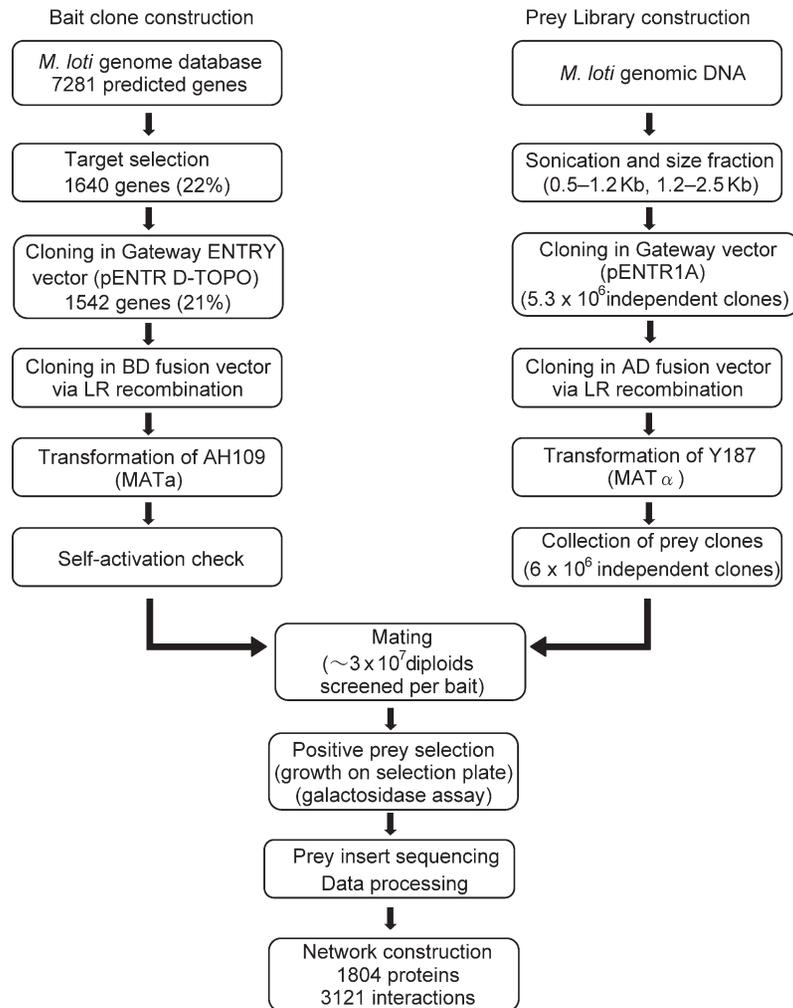
Bait clone construction

- *M. loti* genome database
  7281 predicted genes
- Target selection
  1640 genes (22%)
- Cloning in Gateway ENTRY vector (pENTR D-TOPO)
  1542 genes (21%)
- Cloning in BD fusion vector via LR recombination
- Transformation of AH109 (MATa)
- Self-activation check

Prey Library construction

- *M. loti* genomic DNA
- Sonication and size fraction (0.5–1.2 Kb, 1.2–2.5 Kb)
- Cloning in Gateway vector (pENTR1A)
  (5.3×10^6 independent clones)
- Cloning in AD fusion vector via LR recombination
- Transformation of Y187 (MAT α)
- Collection of prey clones
  (6×10^6 independent clones)

- Mating
  ($\sim 3 \times 10^7$ diploids screened per bait)
- Positive prey selection
  (growth on selection plate)
  (galactosidase assay)
- Prey insert sequencing
  Data processing
- Network construction
  1804 proteins
  3121 interactions

Figure 1. Flow chart of the sequential steps in the YTH analysis of *M. loti* (see Materials and Methods for details). BD and AD indicate the GAL4 DNA-binding domain and GAL4 activation domain, respectively.

three rhizobium species, *M. loti* strain MAFF303099 (symbiont for *Lotus japonicus*), *Sinorhizobium meliloti* strain 1021 (alfalfa symbiont), and *Bradyrhizobium japonicum* strain USDA110 (soybean symbiont), whose complete genome sequences were available when this study was initiated. When genes were selected by the criteria described in Materials and methods, 2797 genes were found to be conserved among three rhizobia species. Of these, we selected 858 *M. loti* genes that were categorized to regulatory function (227 genes), hypothetical protein (479 genes), and unknown protein (152 genes). Altogether, a total of 1640 genes, including 201 genes selected for other purposes, were selected as target genes for this study (Supplementary Table 1).

3.2. Assessment of the PPI data

Of the 1640 genes targeted, 1542 (21% of all predicted *M. loti* genes) bait clones were constructed successfully (Table 2). PPIs were identified on 57% of the bait clones, resulting in 3121 putative interaction pairs consisting of 1804 *M. loti* proteins. (Fig. 2, Supplementary Table 2; all PPI data are also available from RhizoBase, http://bacteria.kazusa.or.jp/rhizobase).

One of the major concerns of large-scale PPI analysis is its reliability. PPI data obtained from comprehensive analyses generally contains numerous false positives, which are mainly caused by promiscuous interactions and self-activation of bait clones.[21,23] Therefore, the detected interactions need to be evaluated by some appropriate criteria. Ito et al. employed interaction sequence tags (ISTs), a pair of tagged sequences obtained from interacting bait and prey clones, to weigh the reliability of each detected interaction.[16] They handled the interactions with high IST hits as 'core' data, which is assumed to be of high relevance and to contain many biologically meaningful interactions. Similarly, as one indicator of data reliability, we classified all detected interactions into

Table 1. *Mesorhizobium loti* genes used for YTH analysis

Description	Number	Screened	Positive	No positive	Self-active
Genes up-regulated in the bacteroid state	93	92	50	41	1
Genes up-regulated under microaerobic condition	72	71	41	29	1
Genes located in symbiosis island	416	391	185	199	7
Genes conserved in other rhizobium species					
Unknown protein	152	130	78	51	1
Hypothetical protein	479	444	255	179	10
Regulatory function	227	220	159	54	7
Others	201	194	117	73	4
Total	1640	1542	885	626	31

four distinct categories (category A–D) based on how many positive clones supported the interaction. Category A and B consist of interactions supported by multiple positive clones with different (A) or identical (B) inserts. Category C consists of interactions supported by a single positive clone. Interactions supported by prey clones that interacted with more than 18 different baits, i.e. 1% of all interacting proteins, were considered promiscuous interactions and classified in category D. Most promiscuously interacting proteins were soluble proteins, and 10 out of 18 promiscuous proteins possessed at least one protein domain known to cause promiscuous interactions[24,25] (Supplementary Table 3). Other than the promiscuous interaction domains, no remarkable physicochemical properties common to the promiscuous proteins, such as isoelectric point or hydrophobicity, were found. The number of interactions in each category is shown in Table 2.

To assess the validity of our categorization, we evaluated all detected interactions using the IG measurement, a method for computationally assessing the reliability of PPI.[21] Interactions with lower IG values are more likely to be reliable than interactions with higher IG values. The IG values for all detected interactions ranged from 1 to 62 and the average IG value was 10.2. When we examined the interactions in which the IG values were <5, $>94\%$ of interactions in category A, B, and C were included, whereas only 32% of category D interactions were included (Supplementary Fig. 1). The average IG values for category A (2.37 ± 0.16), B (2.21 ± 0.17), and C (2.75 ± 0.72) were significantly lower $(P < 0.01)$ than that of category D (26.0 ± 3.49) when 50 independent interactions were selected randomly from each category and compared. This result indicates that our classification is appropriate to define the reliability of each interaction.

To minimize false positives caused by self-activation, we used multiple reporter genes driven by different GAL4-responsive promoters and carefully determined the level of self-activation of each bait clone (see Materials and methods). Furthermore, we re-screened bait proteins under more stringent conditions when the bait protein generated an excessive number of positive colonies (>300 colonies per bait). As shown in Table 2, 31 bait proteins displayed strong self-activation which could not be suppressed by leaky HIS3 reporter gene expression even in the presence of 50 mM 3-AT and the interactions derived from these strong self-active bait clones were discarded. A total of 148 bait clones (9.6% of all screened baits) were screened in the presence of 50 mM 3-AT (Supplementary Table 4). As expected, several bait clones harboring genes with regulatory functions such as transcriptional regulators or two-component response regulators (RRs) showed self-activation. However, 92 out of 148 were genes annotated as hypothetical or unknown protein. To examine the protein domains responsible for their self-activation, we examined the known protein domains assigned in these genes. Although several kinds of bacterial regulatory protein domains, such as LuxR (IPR000792), LysR (IPR000847), and RR

Table 2. Summary of experimental results

| Description	Number
Mesorhizobium loti genes targeted	1640
Mesorhizobium loti genes screened	1542
Number of prey clones assessed per bait	$\sim 3 \times 10^7$
Genes identified to have interactions	1804
Genes showing strong self-activation	31
Total number of positive prey clones collected	13 260
Total identified protein pairs	3121
Interaction supported by multiple positive clones with different inserts (category A)	200
Interaction supported by multiple positive clones with the same inserts (category B)	174
Interaction supported by a single positive clone (category C)	1655
Interaction of putative promiscuous prey clones (category D)	1092
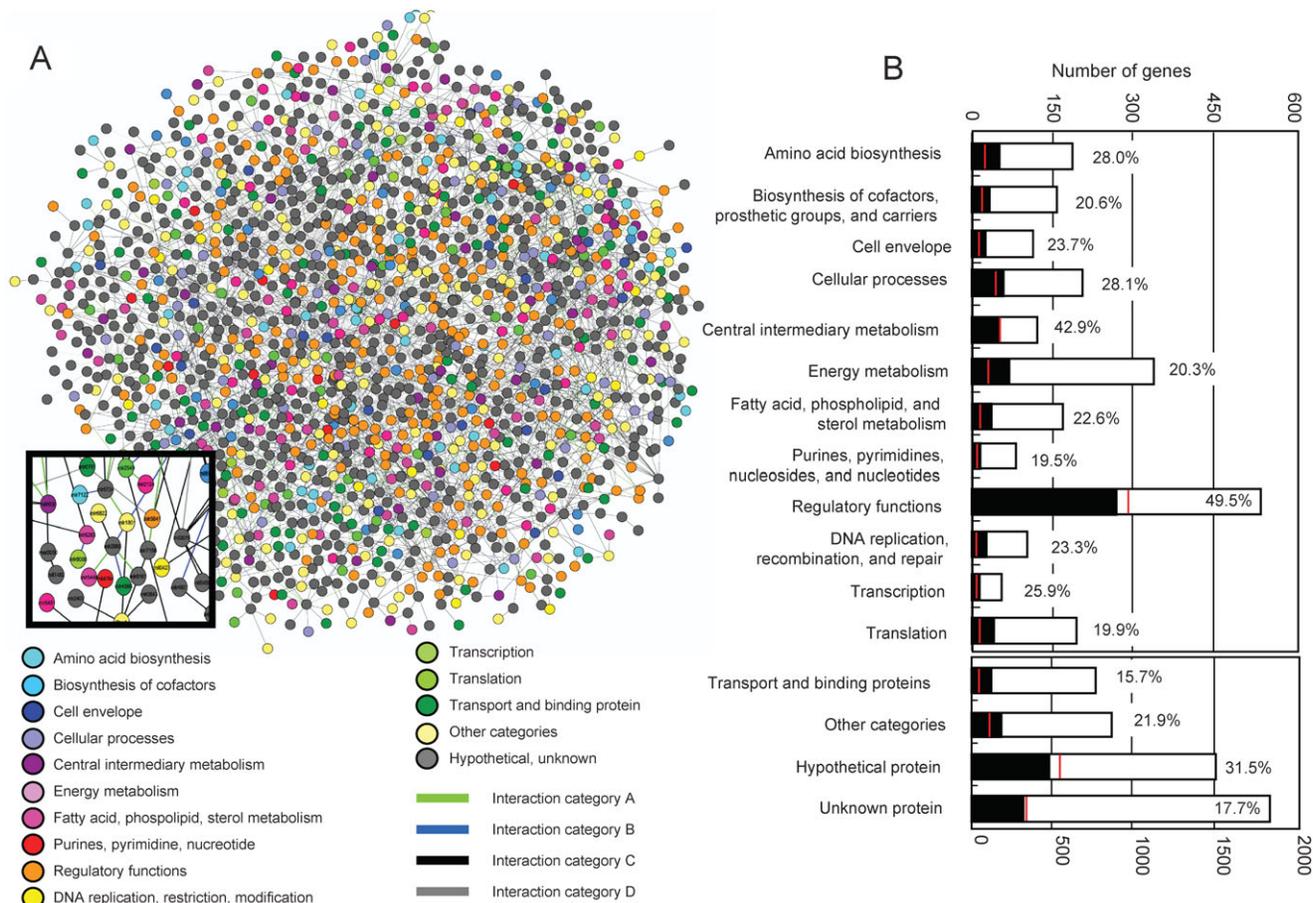
Figure 2. Global view of the PPIs of *M. loti*. (**A**) All detected PPIs. Proteins (circles) are color-coded according to their functional category assigned by Kaneko et al.[1] Interactions (lines) are colored according to the interaction category (A–D), which is classified based on their frequency of detection of identical pairs. (**B**) Number of identified interactions in each function category. The white bar indicates the total number of genes in each functional category assigned in the *M. loti* genome and the black bar indicates the number of genes shown to have interactions. The red bar indicates the number of genes used as bait in the screening. Percentages represent the coverage of interacting proteins in each function category.

receiver (REC) (IPR001789) domain, are frequently found among self-active genes, some proteins with a domain of unknown function (DUF domain) and hypothetical or unknown proteins with no known protein domains were also included.

In addition to the detection of false positives, large-scale PPI analyses tend to miss a large number of known interactions (false negatives). The precise proportion of false-negative interactions can be determined by comparison with published experimental PPI data.[18] However, this approach is difficult to apply to *M. loti* PPI analysis because the available data regarding protein interactions of *M. loti* are limited. To compensate for this problem, we estimated the proportion of false-negatives based on the interactions of proteins in two-component signal transducers whose interaction can be predicted by the location of the corresponding genes in the genome. In the *M. loti* genome, 30 pairs of genes encoding a sensor histidine kinase (HK) and RR are considered to forms operons and, of these, 15 pairs

were detected from screening all RR and 37 HK used as bait, indicating that ∼50% of the interactions could not be detected by our analysis.

3.3. General features of PPIs

A global view of the protein interaction network is illustrated in Fig. 2A. Proteins (nodes) and interactions (edges) are colored according to their functional categories and interaction categories, respectively. Proteins in the obtained protein network cover all functional categories of *M. loti* (Fig. 2B), and interactions are presented as scale-free network, since most proteins had few connections and only a small number had many connections.

Prokaryotic genes are generally organized into operons in which genes are transcribed as a polycistronic mRNA. Several previous studies have demonstrated that genes encoded in the same operon are likely to be coordinately linked and carry out related functions.[26,27] In all the interactions detected, we

found 36 interacting pairs that were encoded by genes that mapped to adjacent loci in the *M. loti* genome (Supplementary Table 5a and b). Notably, 14 out of 36 protein pairs were category A interactions and the frequency of category A interactions between proteins encoded by adjacent genes is significantly higher than that of the whole network. This result further supports the validity of our evaluation of protein interaction data by interaction category. Among the interacting protein encoded at adjacent loci, 15 interactions contained at least one protein of unknown function. Among these, we found a putative part of a sarcosine oxidase complex composed of Mll6238 (annotated as sarcosine oxidase alpha subunit) and Mll6237 (unknown protein that contains a protein domain of sarcosine oxidase gamma subunit). We also found a putative protein complex required for chromosome condensation and segregation between Mll1088 (hypothetical protein contains ScpA domain; IPR003768) and Mll1087 (hypothetical protein contains ScpB domain; IPR005234). Considering that genes belonging to the same operon are predicted to have related functions, the interactions extracted according to the genomic position permit us to predict the biological function of unknown proteins and provides experimental evidence to reinforce the functional relationships between proteins encoded by adjacent genes.

We also identified 25 self-interacting proteins and 9 hetero-dimeric interactions that occurred between two paralogous proteins (Supplementary Table 6). Approximately half of the self-interacting proteins have regulatory functions, and some contained several types of helix-turn-helix (HTH) motifs. This result is reasonable because HTH-containing proteins are known to execute their function in the form of homo-dimers or homo-tetramors[28] and homo-dimeric forms of these proteins have been identified in several microorganisms by large-scale PPI analysis and X-ray diffraction.[29,30] Furthermore, as observed in yeast and some other eucaryotes,[31] the paralogous-interacting proteins detected here tended to exhibit self-interaction. For example, the self-interacting Mll3429 and Mlr5643 also interacted with their paralogues, Mll2335 and Mll3718, respectively. In addition, three other proteins (Mlr6361, Shikimate kinase; Mll8202, GroES; and Mlr2806, NolR) are also known to form homo-dimers,[32–34] indicating that this tendency may support the hypothesis that duplication of self-interacting proteins can generate paralogous proteins whose interactions create functional and structural diversification.[22,31]

3.4.   Interaction of symbiosis related proteins

YTH screening with 581 genes that were selected based on their expression profiles and location in symbiosis island generated 646 interaction pairs (excluding category D interactions). Among the relatively reliable interactions (interaction categories A and B), the proportion of interactions between proteins encoded in symbiosis island was significantly higher (20%) than the ratio expected (8%) by random interactions. This result may reflect the existence of interrelated functions among proteins within the symbiosis island, as suggested by the clustered expression of genes in symbiosis island under the bacteroid state.[14]

In the *M. loti* genome, 40 genes for nodulation and 46 genes for nitrogen fixation were assigned by whole genome sequencing.[1] Interactions were detected on 13 genes for nodulation and 27 genes for nitrogen fixation, respectively. Among these interactions, we successfully identified several interaction pairs whose functional relationships have been proven experimentally, such as NtrX (Mlr0400)−NtrY (Mlr0399) and FixL (Mll6607)−FixJ (Mll6606), which are two-component sensor-regulator pairs that participate in nitrogen metabolisms and sensing of environmental oxygen tension,[35] and NtrR (Mll1670)−NtrP (Mll1671) operon that function in complex as toxin-antitoxin module.[36] Detection of these known interactions indicates that the rhizobial proteins we screened here retained their native conformation in yeast cells.

One of the advantages of interactome analysis is that interactions between novel proteins and well-characterized ones provide us highly informative hints to expand our knowledge, and such interactions were found on several symbiosis-related proteins. For example, a two-component RR (Mll9592) encoded in the *M. loti* plasmid (pMLb) interacted with two distinct NifA proteins (Mll5857 and Mll5837). NifA is a transcriptional activator that controls, in concert with RNA polymerase sigma factor, the expression of genes for nitrogen fixation.[37,38] Unlike other known sigma factor-interacting transcriptional activators such as DctD and NtrC,[39] NifA protein lacks the REC domain which accept the phosphoryl signal from the cognate sensor HK. In contrast, Mll9592 contains a REC domain but lacks any DNA binding domains. Considering their physical interaction and the distribution of protein domains, Mll9592 may execute its function by forming a complex with NifA proteins.

We identified several interactions of proteins encoded by genes whose expression is up-regulated in the bacteroid state or by microaerobic condition. It is noteworthy that approximately one-third of the proteins that interacted with proteins encoded by the up-regulated genes are functionally unannotated proteins. For example, msr6604 which encodes small protein of unknown function that located upstream of the *FixLJ* (mll6606−mll6607) operon interacted with both FixL (Mll6607) and FixJ (Mll6606) (Supplementary

Table 7). Likewise, Mll9215 which was annotated as unknown protein interacted with two distinct FixO proteins (Mll6629 and Mlr6412) (Supplementary Table 7) and expression of the three genes encoding these proteins was up-regulated in the bacteroid state.[14] Furthermore, interaction of proteins encoded by up-regulated genes also revealed putative interaction on several molecular chaperones. For instance, Mll3429 (endopeptidase Clp ATP-binding chain B; ClpB) interacted with Mll2335 (probable ClpA/B-type protease). The expression of mll3429 was up-regulated under microaerobiosis[14] and its orthologous protein (blr1404) of *B. japonicum* USDA110 were identified to be expressed specifically in soybean nodule.[9,13] This result indicates that interaction between Mll3429 and Mll2335 may be a part of protease complex that function in protein processing during symbiosis. In addition, Mll3623 (unknown protein) interacted with two paralogous heat-shock proteins (Mlr4721 and Mlr4720) encoded by genes in the same operon. The genes encoding these heat-shock proteins were up-regulated under microaerobic condition,[14] indicating that Mll3623 may be a target of these heat-shock proteins or may be involved in protection of proteins against microaerobic stress-induced protein denaturation and aggregation. Although the biological significance of these interactions remains to be solved, our results imply that many proteins of unknown function are involved in various aspects of symbiotic nitrogen fixation and our PPI data will provide useful clues to reveal the functional relationships between them.

3.5. Interactions of proteins with regulatory functions

In order to acquire PPI information leads to understanding of signal transduction pathways that functioned in rhizobium, we screened *M. loti* genes with regulatory functions. Among 295 selected genes (including 234 conserved genes), 286 bait clones were used for screening and 618 PPIs were obtained from 207 bait clones.

Among these PPIs, many interaction pairs of two-component signal transducers were detected as reliable interactions (Fig. 3). Two-component signal transduction systems, composed of a HK and RR, are the predominant systems by which bacteria sense several environmental changes, and through a linear phosphorelay from the HK to its cognate RR, cells can adapt rapidly to new conditions. Considering their specific functional relationship, the interaction of two-component signal transducers are suitable models not only to evaluate our PPI data but also to reveal novel part of signaling pathways in *M. loti*. In the *M. loti* genome, 46 and 58 genes encoding HK and RR are assigned, respectively, and, among them,

30 pairs of HK and RR are considered to be transcribed in single operons.[40] From screening of 38 HKs and all the RRs as bait, 33 interactions between HK and RR were obtained (Fig. 3). Among these interactions, we successfully identified 15 pairs of HK and RR that encoded by the same operon, and most of these interactions were supported by multiple positive clones. These interactions, for examples, identified a part of putative chemotaxis pathway of *M. loti* (CheW; mll9513−CheA; mll9511−CheA; mll9511−CheY; mll9509) as observed in PPI analysis of other bacteria.[30,41] Detection of putative cognate pairs of HK and RR as relatively reliable interactions strongly supports the validity of the screen scale and the specificity of our YTH analysis. We also identified 18 pairs of interactions between HK and RR that were located in different regions of the *M. loti* genome (e.g. Mll7700−Mll0861). Since it is difficult to know the specific functional partner of HK or RR from their sequence, the interaction pairs described here provide evidence to support the functional relationships among putative cognate HK and RR encoded in a single operon and also allow us to predict the functional partners of HK or RR located in different regions of the *M. loti* genome. In addition to one-to-one interactions between HK and RR, we found several interactions that occurred between multiple HKs and a single RR, and vice versa (e.g. Mll6691 and Mlr6540). The existence of cross-regulation in two-component systems has been reported in several bacteria,[42] but not in rhizobia. Therefore, the PPIs obtained here for multiple HKs and RRs reveals, for the first time, putative cross-regulation in two-component systems in rhizobia.

3.6. Interactions of proteins of unknown function

Among the 7281 predicted ORFs of *M. loti*, ∼46% (3371 ORFs) are functionally unannotated (categorized as hypothetical or unknown protein).[1] To assign some functional information to hypothetical and unknown proteins of *M. loti*, we examine all the interactions containing these proteins and characterized them based on their interactions with partners of known function. When 877 genes of unknown function were screened as bait (including 608 genes screened as conserved genes), 1598 interactions between a protein of known function and unknown function were obtained (Table 3). Of these, 569 proteins of unknown function had at least one partner of known function and 94 showed interactions with two or more known protein of same function category. For example, Mlr0746 (unknown protein) interacted with three distinct transcriptional regulators (Mlr0745, Mll5360, and Mll2255) which are paralogous proteins with the same protein domains

Figure 3. Interaction pairs of two-component signal transducers. Sensor HK and RR are shown by the blue and the orange boxes, respectively. Boxes marked with a red line indicate the interactions between HK and RR that are encoded by the same operon. The arrow in each interaction indicates the direction of bait protein to prey protein and the reliability of each interaction. HK and RR are designated according to the classification described in Hagiwara et al.[38] hHK, hybrid sensor HK; CheA, CheA-type HK; RR(c), NtrC-family RR; RR(l), NarL-family RR; RR(r), OmpR-family RR; RR(y), CheY-family RR; RR(y), unclassified RR.

(Supplementary Table 7), suggesting that Mlr0746 may have any roles in regulating transcription. Furthermore, to obtain more detailed information about the function unknown proteins, we examined the PPIs of function unknown proteins at the level of protein domains. By focusing on common interaction partners or common protein domains, we assigned 42 unannotated (including 25 conserved) proteins to nine distinct functional categories (Supplementary Table 7). This approach allowed us to identify

Table 3. Summary of interactions with proteins of unknown function

Description	Number
Number of assessed baits of unknown proteins	877
Number of baits of function unknown protein exhibiting interactions	453
Number of function unknown proteins in the entire network	808
Number of function unknown proteins interact with annotated proteins	569
Number of function unknown proteins interact with more than two annotated proteins of the same function category	94
Number of function unknown proteins interact with a common protein of known function or more than two proteins with common domains	42

component of protein complexes that had not been assigned by gene annotation. For example, Mll2736 (hypothetical protein) which contains a ClpS core domain (IPR003769) interacts with two distinct Clp proteases (Mll0663 and Mll2335). Mlr3346 (hypothetical protein) contains a phosphonate metabolism PhnJ domain (IPR010306) and interacts with paralogous hypothetical proteins (Mll9155 and Mlr3342) that contain a phnG domain (IPR009609). On the basis of the fact that proteins with related functions tend to interact, these interactions should reflect the functional properties of the unannotated proteins of *M. loti*. Actually, previous works have predicted the functions of unannotated proteins from large-scale PPI data using similar strategies.[43-45] Since the functional relationships described here could not be determined from genome sequences or other genome-wide analyses, our data should provide novel and useful information for elucidating the biological roles of many unannotated proteins of rhizobia.

Our results provide a comprehensive data source for PPIs of *M. loti* proteins. From this PPI data, we have predicted putative novel relationships among proteins for symbiotic nitrogen fixation and signal transduction and have provided some functional information on several unannotated proteins. Since our study examined primarily on proteins playing certain roles in symbiotic nitrogen fixation (i.e. one of the unique characteristics of rhizobium) and proteins conserved in other rhizobia, the data are applicable to further functional studies of many rhizobial species, as well as *M. loti*. To make the interaction data publicly available, we provide all obtained PPI data through rhizobial genome database, RhizoBase (http://bacteria.kazusa.or.jp/rhizobase/index.html). In the database, we have provided all PPI data including the interaction category and interacting region of the prey protein so that users can obtain PPI information depending on their own needs.

Although some of our findings will be needed to be confirmed by independent approaches, we believe that the obtained PPI data should provide a useful starting point to elucidate the biological function of many rhizobial genes.

References

1. Kaneko, T., Nakamura, Y., Sato, S., et al. 2000, Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*, *DNA Res.*, **31**, 331–338.
2. Kaneko, T., Nakamura, Y., Sato, S., et al. 2002, Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110, *DNA Res.*, **9**, 189–197.
3. Galibert, F., Finan, T. M., Long, S. R., et al. 2001, The composite genome of the legume symbiont *Sinorhizobium meliloti*, *Science*, **293**, 668–672.
4. González, V., Santamaría, R. I., Bustos, P., et al. 2006, The partitioned *Rhizobium etli* genome: genetic and metabolic redundancy in seven interacting replicons, *Proc. Natl. Acad. Sci. USA*, **103**, 3834–3839.
5. Young, J. P., Crossman, L. C., Johnston, A.W., et al. 2006, The genome of *Rhizobium leguminosarum* has recognizable core and accessory components, *Genome Biol.*, **7**, R34.
6. Giraud, E., Moulin, L., Vallenet, D., Barbe, V., et al. 2007, Legumes symbioses: absence of Nod genes in photosynthetic bradyrhizobia, *Science*, **316**, 1307–1312.
7. Van Sluys, M. A., Monteiro-Vitorello, C. B., Camargo, L. E. A., et al. 2002, Comparative genomic analysis of plant-associated bacteria, *Ann. Rev. Phytopathol.*, **40**, 169–189.
8. Noel, K. D., Sanchez, A., Fernandez, L., Leemans, J. and Cevallos, M. A. 1984, Rhizobium phaseoli symbiotic mutants with transposon Tn5 insertions, *J. Bacteriol.*, **158**, 148–155.
9. Hoa, L. T. P., Nomura, M. and Tajima, S. 2004, Characterization of Bacteroid Proteins in Soybean Nodules Formed with *Bradyrhizobium japonicum* USDA110, *Microb. Environ.*, **19**, 71–75.
10. Bergès, H., Lauber, E., Liebe, C., et al. 2003, Development of *Sinorhizobium meliloti* Pilot Macroarrays for Transcriptome Analysis, *Appl. Environ. Microbiol.*, **69**, 1214–1219.
11. Djordjevic, M. A., Chen, H. C., Natera, S., et al. 2003, A global analysis of protein expression profiles in *Sinorhizobium meliloti*: discovery of new genes for nodule occupancy and stress adaptation, *Mol. Plant Microbe Interact.*, **16**, 508–524.
12. Barnett, M. J., Toman, C. J., Fisher, R. F. and Long, S. R. 2004, A dual-genome Symbiosis Chip for coordinate

study of signal exchange and development in a prokaryote−host interaction, *Proc. Natl. Acad. Sci. USA*, **101**, 16636−16641.

13. Sarma, A. D. and Emerich, D. W. 2005, Global protein expression pattern of *Bradyrhizobium japonicum* bacteroids: a prelude to functional proteomics, *Proteomics*, **5**, 4170−4184.

14. Uchiumi, T., Oowada, T., Itakura, M., et al. 2004, Expression islands clustered on symbiosis island of *Mesorhizobium loti* genome, *J. Bacteriol.*, **45**, 2439−2448.

15. Uetz, P., Giot, L., Cagney, G., et al. 2000, A comprehensive analysis of protein−protein interactions in *Saccharomyces cerevisiae*, *Nature*, **403**, 623−627.

16. Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M. and Sakaki, Y. 2001, A comprehensive two-hybrid analysis to explore the yeast protein interactome, *Proc. Natl. Acad. Sci. USA*, **98**, 4569−4574.

17. Walhout, A. J. M., Sordella, R., Lu, X., et al. 2000, Protein interaction mapping in *C. elegans* using proteins involved in vulval development, *Science*, **287**, 116−121.

18. Giot, L., Bader, J. S., Brouwer, C., et al. 2003, A protein interaction map of *Drosophila melanogaster*, *Science*, **302**, 1727−1736.

19. Nakayama, M., Kikuno, R. and Ohara, O. 2002, Protein−protein interactions between large proteins: two-hybrid screening using a functionally classified library composed of long cDNAs, *Genome Res.*, **12**, 1773−1784.

20. Chen, W. P. and Kuo, T. T. 1993, A simple and rapid method for the preparation of gram-negative bacterial genomic DNA, *Nucleic Acid Res.*, **21**, 2260.

21. Saito, R., Suzuki, H. and Hayashizaki, Y. 2002, Interaction generality, a measurement to assess the reliability of a protein−protein interaction, *Nucleic Acids Res.*, **30**, 1163−1168.

22. Sato, S., Shimoda, Y., Muraki, A., Kohara, M., Nakamura, Y. and Tabata, S. 2007, A large scale protein−protein interaction analysis in *Synechocystis* sp. PCC6803, *DNA Res.*, in press.

23. Suzuki, H., Fukunishi, Y., Kagawa, I., et al. 2007, Protein−protein interaction panel using mouse full-length cDNAs, *Genome Res.*, **11**, 1758−1765.

24. Ng, S-K., Zhang, Z., Tan, S-H. and Lin, K. 2003, InterDom: a database of putative interacting protein domains for validating predicted protein interactions and complexes, *Nucleic Acids Res.*, **31**, 251−254.

25. Arifuzzaman, M., Maeda, M., Itoh, A., et al. 2006, Large-scale indentification of protein−protein interaction of *Escherichia coli* K-12, *Genome Res.*, **16**, 686−691.

26. Overbeek, R., Fonstein, M., D'Souza, M., Pusch, G. D. and Maltsev, N. 1999, The use of gene clusters to infer functional coupling, *Proc. Natl. Acad. Sci. USA*, **96**, 2896−2901.

27. Snel, B., Bork, P. and Huynen, M. A. 2002, The identification of functional modules from the genomic association of genes, *Proc. Natl. Acad. Sci. USA*, **99**, 5890−5895.

28. Rosinski, J. A. and Atchley, W. R. 1999, Molecular evolution of helix-turn-helix proteins, *J. Mol. Evol.*, **49**, 301−309.

29. Hall, D. R., Gourley, D. G., Leonard, G. A., et al. 1999, The high-resolution crystal structure of the molybdate-

dependent transcriptional regulator (ModE) from *Escherichia coli*: a novel combination of domain folds, *EMBO J.*, **18**, 1435−1446.

30. Usui, K., Katayama, S., Kanamori-Katayama, M., et al. 2005, Protein−protein interactions of hyperthermophilic archeaon *Pyrococcus horikoshii* OT3, *Genome Biol.*, **5**, R98.

31. Pereira-Leal, J. B., Levy, E. D., Kamp, C. and Teichmann, S. A. 2007, Evolution of protein complexes by duplication of homomeric interactions, *Genome Biol.*, **8**, R51.

32. Romanowski, M. J. and Burle, S. K. 2002, Crystal structure of the *Escherichia coli* shikimate kinase I (AroK) that confers sensitivity to mecillinam, *Proteins*, **47**, 558−562.

33. Xu, Z., Horwich, A. L. and Sigler, P. B. 1997, The crystal structure of the asymmetric GroEL-GroES-(ADP)7 chaperonin complex, *Nature*, **388**, 741−750.

34. Cren, M., Kodorosi, A. and Kndorosi, E. 1994, An insertional point mutation inactivates NolR repressor in *Rhizobium meliloti* 1021, *J. Bacteriol.*, **176**, 518−519.

35. Gilles-Gonzalez, M. A., Ditta, D. S. and Helinski, D. R. 1991, A hemoprotein with kinase activity encoded by the oxygen sensor of *Rhizobium meliloti*, *Nature*, **350**, 170−172.

36. Bodogai, M., Ferenczi, S., Bashtovyy, D., Miclea, P., Papp, P. and Dusha, I. 2006, The ntrPR operon of *Sinorhizobium meliloti* is organized and functions as a toxin-antitoxin module, *Mol. Plant Microbe Interact.*, **19**, 811−822.

37. Labes, M., Rastogi, V., Watson, R. and Finan, T. M. 1993, Symbiotic nitrogen fixation by a nifA deletion mutant of *Rhizobium meliloti*: the role of an unusual ntrC allele, *J. Bacteriol.*, **175**, 2662−2673.

38. Martinez-Argudo, I., Little, R., Shearer, N., Johnson, P. and Dixon, R. 2004, The NifL-NifA system: a multidomain transcriptional regulatory complex that integrates environmental signals, *J. Bacteriol.*, **186**, 601−610.

39. Labes, M. and Finan, T. M. 1993, Negative regulation of sigma 54-dependent dctA expression by the transcriptional activator DctD, *J Bacteriol.*, **175**, 2674−2681.

40. Hagiwara, D., Yamashino, T. and Mizuno, T. 2004, Genome-wide comparison of the His-to-Asp phosphorelay signaling components of three symbiotic genera of Rhizobia, *DNA Res.*, **11**, 57−65.

41. Rain, J. C., Selig, L., De Reuse, H., et al. 2001, The protein−protein interaction map of *Helicobacter pylori*, *Nature*, **409**, 211−215.

42. Verhamme, D. T., Arents, J. C., Postma, P. W., Crielaard, W. and Hellingwerf, K. J. 2002, Investigation of *in vivo* cross-talk between key two-component systems of *Escherichia coli*, *Microbiology*, **148**, 69−78.

43. Schwikowski, B., Uetz, P. and Fields, S. 2000, A network of protein−protein interactions in yeast, *Nat. Biotechnol.*, **18**, 1257−1261.

44. Samanta, M. P. and Liand, S. 2003, Predicting protein functions from redundancies in large-scale protein interaction networks, *Proc. Natl. Acad. Sci. USA*, **100**, 12579−12583.

45. Okada, K., Kanaya, S. and Asai, K. 2005, Accurate extraction of functional associations between proteins based on common interaction partners and common domains, *Bioinformatics*, **21**, 2043−2048.