TITLE

Massive rhizobial genomic variations associated with partner quality in *Lotus*–*Mesorhizobium* symbiosis

AUTHORS

Masaru Bamba[1*], Seishiro Aoki[2], Tadashi Kajita[3], Hiroaki Setoguchi[4], Yasuyuki Watano[5], Shusei Sato[6], and Takashi Tsuchimatsu[5*]

AFFILIATIONS AND ADDRESSES

[1] Department of Biology (Frontier Science Program), Graduate School of Science and Engineering, Chiba University, 1-33 Yayoi, Inage, Chiba 263-8522, Japan

[2] Department of Biological Sciences, Graduate School of Science, The University of Tokyo, 2-11-16 Yayoi, Bunkyo-ku, Tokyo 113-0032, Japan

[3] Iriomote Station, Tropical Biosphere Research Center, the University of Ryukyus, 870 Uehara, Taketomi-cho, Yaeyama-gun, Okinawa 907-1541, Japan

[4] Graduate School of Human and Environmental Studies, Kyoto University, Yoshidanihonmatsu-cho, Sakyo-ku, Kyoto 606-8501 Japan

[5] Department of Biology, Graduate School of Science, Chiba University, 1-33 Yayoi, Inage, Chiba 263-8522, Japan

[6] Graduate School of Life Sciences, Tohoku University, 2-1-1 Katahira, Aoba, Sendai 980-8577, Japan

CORRESPONDING AUTHORS

Takashi Tsuchimatsu: takashi@chiba-u.jp

Masaru Bamba: contact2093@gmail.com

KEYWORDS

Legume–rhizobia mutualism, whole-genome sequencing, partner quality variation, horizontal gene transfer

ABSTRACT

In diverse mutualistic relationships, genetic variations in impact on the growth of interacting partners—variations in partner quality—are common, despite the theoretical prediction that selection favoring high-quality partners should eliminate such variations. Here, we investigated how variations in partner quality could be maintained in the nitrogen-fixing mutualism between *Lotus japonicus* and *Mesorhizobium* bacteria. We reconstructed *de novo* assembled full-genome sequences from nine rhizobial symbionts, finding massive variations in the core genome and the contrastingly similar symbiotic islands, indicating recent horizontal gene transfer (HGT) of the symbiosis islands into diverse *Mesorhizobium* lineages. A cross-inoculation experiment using nine sequenced rhizobial symbionts and 15 *L. japonicus* accessions revealed extensive quality variations represented by plant growth phenotypes, including genotype-by-genotype interactions. Quality variations were not associated with the presence/absence variations of known symbiosis-related genes in the symbiosis island, but rather, showed significant correlations with the core genome variations, supported by SNP- and kinship matrix-based association analyses. These findings highlight the novel role of HGT of symbiosis islands, which indirectly supply mutations of core genomes into *L. japonicus*-associated bacteria, thereby contributing to the maintenance of variations in partner quality.

INTRODUCTION

While mutualistic symbiotic relationships—unlike organisms living together and establishing a cooperative interaction—are ubiquitous in nature, it remains unclear how mutualistic interactions originated and how they are evolutionarily maintained. Evolutionary theory predicts that mutualism might be unstable because natural selection would favor mutualists that optimize their fitness by minimizing the costs of returning benefits to a partner (Frederickson 2013; Ghoul *et al.* 2013). Therefore, the stabilizing mechanisms that prevent the invasion of low-quality partners have been a major focus of studies (Heath and Stinchcombe 2014), such as those involving partner choice (Bull and Rice 1991), partner fidelity feedback (Bull and Rice 1991; Archetti *et al.* 2011) and sanctions against such invasion (Kiers *et al.* 2003).

Although such stabilizing mechanisms should reduce variations in partner quality in mutualistic symbiosis, thereby maintaining high-quality mutualists, variations in partner quality have been observed in diverse mutualistic relationships (Thrall *et al.* 2000; Sachs *et al.* 2010;). Multiple models have been proposed to resolve this discrepancy, such as mutation–selection balance and spatiotemporally varying selection (Van Dyken *et al.* 2011; Simonsen and Stinchcombe 2014; Smith *et al.* 2014; Steidinger and Bever 2014; reviewed by Heath and Stinchcombe 2014). The mutation–selection balance is a model postulating that low-quality partners evolve via mutations but are slowly purged from populations by purifying selection. The model of spatiotemporally varying selection predicts that the fitness of a partner's genotype varies depending on the genotype of an interacting species, such as genotype × genotype (G × G) interactions, on the spatially variable environment conditions (G × E interactions) or on temporally variable selection (Denison and Kiers 2004). While these models provide

72 possible explanations for the persistence of variations in partner quality in mutualist
73 relationships, it remains unclear how variations in partner quality arise and how they are
74 maintained in natural populations.

75 The legume–rhizobia mutualism is an ideal model to address this question
76 because we can manipulate genotypes of both species and reconstruct their interactions
77 *in vitro*. To understand the origin and the maintenance of variations in partner quality at
78 the microevolutionary scale, it is essential to quantify the quality variations among
79 strains originated from single leguminous species and disentangle the genetic basis
80 underlying the rhizobial quality. Although variations in rhizobial quality have been
81 observed in multiple systems of legume–rhizobia mutualisms (e.g. *Acasia–Ensifer*,
82 Barrett *et al.* 2015 and 2016; *Medicago–Ensifer*, Porter *et al.* 2011), most previous
83 studies were based on a limited number of genes, and thus, the genetic basis underlying
84 the rhizobial quality was unclear. Although Porter *et al.* (2019) and Klinger *et al.* (2016)
85 analyzed the genetic basis of the rhizobial quality variation using genome-wide
86 polymorphism data, these studies used a limited number of plant strains, and therefore,
87 did not take G × G interactions into account. Furthermore, genome-wide polymorphism
88 data of these previous studies were based on resequencing using Illumina short-reads,
89 making it difficult to investigate structural variations suggested to be important for
90 plant–microbe interactions (Raffaele *et al.* 2010; Tsushima *et al.* 2019).

91 Here, we focused on the mutualism between *Lotus japonicus* (Regel) K. Larsen
92 and their rhizobial symbionts. *Lotus japonicus* has been regarded as a model species for
93 the understanding of plant–microbe interactions (Bamba *et al.* 2019a), and there have
94 been extensive studies on the molecular, physiological and genomic bases of plant–
95 rhizobia symbiosis (Handberg and Stougaard 1992; Szczyglowski *et al.* 1998; Kouchi *et*

al. 2004; Maekawa *et al.* 2009; Madsen *et al.* 2010; Suzuki *et al.* 2011; Soyano *et al.* 2013; Nishida *et al.* 2016, 2018). Among *L. japonicus*-associated symbionts, full-genome sequence information is available from two strains, *Mesorhizobium japonicum* MAFF303099 and *M. loti* TONO (Kaneko *et al.* 2000; Shimoda *et al.* 2016). Bamba *et al.* (2019b) explored the genetic diversity of *L. japonicus*-associated symbionts, finding that *L. japonicus* in natural populations were associated with highly diverse *Mesorhizobium* bacteria. However, they used only three housekeeping and five symbiotic genes, and the detailed genomic variations of *L. japonicus*-associated symbionts are still unknown. In this study, to investigate the genomic variations of *L. japonicus*-associated symbionts, we first reconstructed high quality *de novo* assembled genome sequences from nine rhizobial symbionts sampled from three geographically distinct locations in Japan. Second, to quantify the rhizobial variations in partner quality including G × G interactions, we performed a cross-inoculation experiment using nine full-genome sequenced rhizobial symbionts and 15 *L. japonicus* natural accessions. Three of those *L. japonicus* accessions originated from the same locations where the nine rhizobial strains were collected, so we could explore a signature of local adaptation: i.e. native rhizobial genotypes outperform foreign rhizobial genotypes when associated with the host genotypes originating from the same locations. Finally, to infer which genomic regions were responsible for rhizobial variations in partner quality, we performed a series of analyses testing the association between genomic polymorphisms and rhizobial variations in partner quality.

MATERIALS AND METHODS

Bacterial strains

We used nine *L. japonicus*-associated *Mesorhizobium* strains for this study, which were previously referred to as 113-1-1, 113-3-3, 113-3-9, 131-2-1, 131-2-5, 131-3-5, L-2-11, L-8-3 and L-8-10 (Bamba *et al.* 2019b). These nine strains were sampled from three geographically distinct localities, Tottori (113-1-1, 113-3-3 and 113-3-9), Aomori (131-2-1, 131-2-5 and 131-3-5) and Miyakojima (L-2-11, L-8-3 and L-8-10) (Supporting Information Table S1), where *L. japonicus* natural accessions of the Natural BioResource Project also originated (MG50 from Tottori, MG23 from Aomori and MG20 from Miyakojima; Supporting Information Table S2). Details of the sampling localities have been described by Bamba *et al.* (2019b).

DNA extraction and whole-genome sequencing using MinION and Illumina HiSeq

Prior to DNA extraction, we cultured rhizobial strains on a tryptone yeast (TY) agar plate for 4 days at 28°C, and then picked single colonies and incubated them for 3 days at 28°C in liquid TY medium. After incubation, we precipitated the cells by centrifugation at 13,000 *g* for 3 min and rinsed them with sterilized MilliQ water (Millipore Corp., Burlington, MA, USA) twice. The genomic DNA of each rhizobial strain was extracted using a NucleoBond CB20 system (MACHERY-NAGEL GmbH & Co. KG, Düren, Germany), according to the manufacturer's instructions. The quality of genomic DNA was confirmed using agarose gel electrophoresis and a BioSpec-nano system (Shimadzu, Kyoto, Japan).

We performed whole-genome sequencing analyses using Oxford Nanopore Technologies (ONT) MinION and Illumina Hiseq 2500 systems. The library for ONT MinION was prepared using Rapid Barcoding kits (SQK-RBK004). We adjusted all nine libraries to the same concentration, mixed them together and then loaded them onto R9.4 flow cells. The sequencing run was performed twice on a MinION MK1b device following the NC_48h_Sequencing_Run_FLO-MIN106_SQK-RBK004 protocol. The library preparation for Illumina HiSeq and sequencing run were performed by DNAFORM (RIKEN, Yokohama, Japan).

Preprocessing of next-generation sequencing data, *de novo* assembly and annotation

The ONT reads were demultiplexed with Albacore 2.2.2 (https://github.com/Albacore/albacore), and adapter sequences were trimmed with Porechop 0.2.3 (https://github.com/rrwick/Porechop). The quality of demultiplexed reads was calculated with NanoStat 1.1.0 (De Coster *et al.* 2018). The ONT reads were processed with NanoFilt 2.2.0 (De Coster *et al.* 2018) to keep sequences with a q-score > 8, and the first 100 bases were removed to increase sequence quality, with a minimum sequence length of 1 kb.

The overall quality of the HiSeq reads was evaluated using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). After confirming the lack of technical errors in the sequencing, low-quality tails were trimmed from each read using SolexaQA (Cox *et al.* 2010) with a cutoff threshold set at a q-score of 30, and reads shorter than 75 bp were filtered with PRINSEQ++ (Schmieder and Edwards,

2011). The Hiseq-filtered reads with pair-end relationships were repaired with BBtools (https://sourceforge.net/projects/bbmap/).

The hybrid read set (both Illumina and ONT reads) for each isolate was assembled using Unicycler 0.4.0 (Wick *et al.* 2017) in its conservative mode. Unicycler performs the assembly of the Illumina reads with SPAdes 3.12.0 (Bankevich *et al.* 2012), and assembly graph scaffolds were then prepared using ONT reads. Unicycler was used to polish the final assembly of Illumina reads, and Pilon (Walker *et al.* 2014) was applied to reduce the rate of small base-level errors. The resulting assembly graph was visualized using Bandage (Wick *et al.* 2015).

The assembled genomes were annotated with the Rapid Annotation using Subsystem Technology (RAST) annotation server (http://rast.theseed.org/FIG/rast.cgi; Aziz *et al.* 2008; Brettin *et al.* 2015; Overbeek *et al.* 2014). Annotation completeness was assessed using BUSCO v3 (Rhizobiales database, Waterhouse *et al.* 2018).

Comparative genomics

We inferred orthologs from all nine assembled genomes and reference genomes of 15 rhizobial strains (*Mesorhizobium*, *Ensifer*, *Rhizobium*, *Bradyrhizobium*, *Azorhizobium*, *Paraburkholderia* and *Cupriavidus*; Supporting Information Table S3), including both chromosomes and plasmids, using SonicParanoid with the most sensitive parameter settings (Consentino and Iwasaki 2018).

We extracted single-copy orthologs that were found in all nine sequenced strains and two reference strains (*M. japonicus* MAFF303099 and *M. loti* TONO). Orthologous groups were assigned as genes in the core genome or in the symbiosis island based on the genome locations of the reference strain, *M. japonicus*

MAFF303099: coordinates of the symbiosis island were 4,643,427 [*intS*] – 5,255,770 [*trnF*-GAA], and the rest was considered to be the core genome (Kaneko *et al.* 2000). We generated a multiple nucleotide sequence alignment of each single-copy orthologous group using MAFFT 7.245 (Katoh and Standley 2013) with the E-INS-i algorithm, and extracted biallelic single nucleotide polymorphisms (SNPs) from the alignments.

To visualize the genetic variations among the nine sequenced strains, we performed principal component analyses (PCA) based on the following datasets: 1) presence/absence of orthologs, 2) the copy number variation of orthologs, 3) SNPs in the core genome and 4) SNPs in the symbiosis island. PCA was performed using the *prcomp* function implemented in R 3.6.1 (R Core Team (2019); http://www.R-project.org/).

To characterize the genome-wide pattern of polymorphisms in these nine sequenced strains, we calculated nucleotide diversity (π values for synonymous and nonsynonymous sites) and Tajima's *D* for each gene using MEGA-CC (Kumar *et al.* 2012; Tamura *et al.* 2011).

To investigate the phylogenetic relationships of *L. japonicus*-associated symbionts with other nodule bacteria, phylogenetic trees of each single-copy gene were reconstructed using the maximum likelihood method with the program RAxML-NG (Kozlov *et al.* 2019), together with orthologs from reference sequences. Prior to the tree reconstruction, haplotypes were determined for each gene based on nucleotide substitutions and indels. The nucleotide substitution models were selected by the Akaike Information Criterion, as implemented in ModelTest-NG (Darriba *et al.* 2019). The single most likely tree out of 10 search replicates was saved for phylogenetic analyses. The outgroups of each phylogenetic tree were determined using the Graph Splitting

method (Matsui and Iwasaki 2019) based on protein sequences of each gene, under the assumption that an outgroup is the most distant operational taxonomic units from the focal group containing *L. japonicus*-associated symbionts. The sensitivity was set to 2.0 in MMseqs2 (Steinegger and Söding 2017), used in the GraphSplitting method.

We then characterized the topologies of maximum likelihood trees of each single-copy gene based on the following criteria: 1) whether *L. japonicus*-associated symbionts (nine sequenced strains in this study, *M. japonicum* MAFF303099 and *M. loti* TONO) formed a single clade; 2) whether *Lotus*-associated symbionts (*L. japonicus*-associated symbionts, *M. loti* NZP2037 and *M. japonicum* R7A) formed a single clade; and 3) whether *Mesorhizobium* strains (nine sequenced strains in this study and other *Mesorhizobium* strains) formed a single clade. This analysis was performed using in-house scripts written in Python 3.

We also performed synteny analysis using the progressiveMauve 2.4.0 program (Darling *et al.* 2010) to identify any structural variations in the symbiosis islands. The symbiosis islands of each genome were first identified by aligning them with that of *M. japonicum* MAFF303099. We then excised the region and aligned the direction using SnapGene software (GSL Biotech; https://www.snapgene.com).

Cross-inoculation experiments

To quantify the effects of rhizobial symbionts, host plants and their interactions on plant phenotypes, we performed a cross-inoculation experiment using nine sequenced rhizobial strains (Supporting Information Table S1) and 15 *L. japonicus* natural accessions (Supporting Information Table S2), resulting in 135 combinations in total. Seeds of *L. japonicus* accessions were obtained from the Natural BioResource Project.

Prior to inoculation experiments, we prepared inoculant strains in the logarithmic growth phase. We cultured rhizobial strains on a TY agar plate for 4 days at 28°C, and then picked single colonies and precultured them with shaking in 2 mL TY liquid media for 3 days at 28°C. Aliquots (200 μL) of precultured strains were transferred into 50 mL lots of TY liquid medium and cultured with shaking at 28°C for 48 h. The cultured strains were precipitated by centrifugation at 5800 g for 3 min, washed with sterilized water three times and adjusted to 1.0×10^7 cells/mL (based on optical density at 600 nm).

Partly scrubbed *L. japonicus* seeds were surface sterilized by immersion in 2% sodium hypochlorite for 3 min and rinsed three times with sterile distilled water. After overnight imbibition, the swollen seeds were sown onto 0.8% agar plates, incubated in the dark for 3 days at 20°C and then grown at 20°C under 16/8 light/dark conditions for 24 h. The rooting plants were transplanted into Leonard jars (Leonard 1943) filled with 300 mL sterilized vermiculite with 300 mL sterilized nitrogen-free B&D medium (Broughton and Dilworth 1971) and grown at 20°C under the same lighting conditions for 3 days. Finally, we inoculated 20 mL of each concentration-adjusted rhizobial strain into Leonard jars and grew them at 20°C under the same lighting conditions for 21 days. We then harvested whole plant bodies, imaged all individuals with a high resolution scanner and separated them into shoots and roots. The shoots were dried over 48 h at 65°C, and then the dry weights were measured. For root phenotypes, we measured the numbers and areas of nodules from the scanned data. Shoot dry weight (SDW in g), number of nodules (NON), total size of nodule (NOA in mm^2) and nodule size per nodule (NOA/NON ratio defined as NOAi in mm^2) were obtained from all individuals used in the experiment. We repeated all inoculation experiments twice. When we grew

L. japonicus without inoculation, plants did not form any nodules and lost most of the leaves (data not shown).

Data analysis of the inoculation experiments

We performed analysis of variance to test whether genotypes of inoculant symbiont, those of host plants and their interactions (G × G) significantly influenced the four measured phenotypes (SDW, NON, NOA and NOAi). We considered mean values of each rhizobial strain (means of phenotypes of 15 plant accessions) as the rhizobial quality. To quantify G × G interactions, we generated Euclidean distance matrices based on the phenotypic differences between rhizobial strains by using the standardized phenotypic values whose mean values of each rhizobial symbiont were set to 0, thereby controlling for bacterial genetic effects. We used mean phenotype values and Euclidean distance matrices for the following association analyses with the bacterial genome sequences. R v. 3.6.1 was used for analysis of variance (R Core Team 2019), and the Euclidean distance was calculated by using SciPy (Virtanen *et al.* 2019).

To determine the genomic regions associated with rhizobial quality, we first performed an SNP-based association analysis that examined the correlation between genome-wide SNPs and rhizobial quality using a linear model implemented by LIMIX (Lipert *et al.* 2014). We then performed Mantel tests to evaluate whether the genetic distance between rhizobial strains was correlated positively with the difference in the variations of rhizobial quality or those of G × G interactions. We performed these analyses for the core genome and symbiosis islands separately to examine which genomic regions were more strongly associated with variations in rhizobial quality. For the genetic distance, kinship matrices were calculated using mixmogam (Segura *et al.*

282 2012). Mantel tests were conducted using Spearman's rank correlation implemented in

283 the scikit-bio v. 0.5.4 (http://scikit-bio.org/) library of Python.

RESULTS

Bacterial genome sequencing

Nine rhizobial genome sequences were obtained using ONT MinION and Illumina HiSeq sequencing. From MinION sequencing, 1,011,067 reads (mean length 6748 bp) were obtained, covering a total of 6.823 Gb. After pre-processing the MinION sequences, 822,488 reads were allocated into the nine samples, which ranged from 47,958 to 142,669 in read number and 370–1181 Mb in total length (Supporting Information Table S4). From HiSeq sequencing, 32,511,368 pre-processing reads were obtained and allocated into the nine samples, which ranged from 2,432,754 to 4,476,543 in read number and 364–671 Mb in total length. All quality-filtered reads from both the MinION and HiSeq sequences were used for the *de novo* assembly analyses.

De novo assembly and annotation

Nearly complete assembled genomes of all nine rhizobial strains were obtained using Unicycler (Wick *et al.* 2017) by combining MinION long reads and HiSeq short reads. All nine rhizobial strains have large circular genomes, considered as chromosomes (6.652–8.451 Mb; Fig. 1 and Supporting Information Table S1). Five of the nine strains had several shorter plasmid-like circular genomes (56–654 kb), and two of the strains had short fragment sequences, which were presumably contaminants (< 10 kb fragment lengths and not closely related to *Mesorhizobium*). The total genome sizes excluding presumable contaminant fragments ranged from 7.108 to 8.451 Mb. The genome sizes of all assembled genomes were similar to that of the reference strain, *M. japonicum* MAFF303099 (chromosome, 7.036 Mb; pMLa, 0.351 Mb, pMLb, 0.251 Mb; and total 7.596 Mb). All genomes without presumed contaminant fragments were used for the

following analyses. Gene predictions from all assembled genomes were conducted using the RAST server (Overbeek *et al.* 2014). The number of coding sequences was 7173–8377, including six rRNA genes and 50–61 tRNA genes in each genome (Supporting Information Table S1). The number of coding sequences was also similar to that of *M. loti* MAFF303099 (7343 genes). The evaluation of completeness of gene prediction by BUSCO analyses showed that all nine rhizobial strains had no missing BUSCO. A few were fragmented (0.292–1.17% fragmented) and most BUSCOs were complete (98.8–99.7% complete; Supporting Information Table S1). These BUSCO analyses indicated that we had successfully reconstructed nearly complete genomes of nine rhizobial strains.

Ortholog analysis

The newly assembled genomes, including both chromosomes and plasmids, were compared with the reference genomes of 15 rhizobial strains (Supporting Information Table S3). We identified a total of 15,712 orthologous groups (OGs) of genes using SonicParanoid (Consentino and Iwasaki 2018). Among these OGs, 3095 were conserved among 11 *L. japonicus*-associated symbionts (*M. japonicum* MAFF303099, *M. loti* TONO and nine newly assembled genomes), and 3047 OGs were conserved among 13 *Lotus*-associated symbionts (the same 11 *L. japonicus*-associated symbionts and two *L. corniculatus*-associated symbionts, *M. japonicum* R7A and *M. loti* NZP2037).

We obtained 2239 OGs as single-copy orthologs found in all of nine sequenced genomes of symbionts and two reference strains (*M. japonicum* MAFF303099 and *M.*

loti TONO; Supporting Information Table S5). Seventy-six single-copy orthologs were located in symbiosis islands based on the coordinates of *M. japonicum* MAFF303099.

Using all the identified OGs, we investigated the presence/absence of genes present on the symbiosis islands of *M. japonicum* MAFF303099 and reported to be related to symbiotic features such as nitrogen fixation, nodulation factor assembly and protein secretion systems (Porter *et al.* 2019; Souza *et al.* 2012; Wang *et al.* 2014). All 28 nitrogen fixation-related genes (*nif*, *fdx* and *fix*) were present in the nine newly assembled genomes and in other genomes of *Lotus*-associated symbionts (Fig. 2). Sixteen nodulation factor assembly genes (*nod*, *nol* and *noe* genes) were present in all nine genomes. All 14 type III protein secretion system (T3SS) genes were absent in one *L. japonicus*-associated strain (131-3-5) and in two *L. corniculatus*-associated strains; two of them (*nolB* and *nolU*) were also missing in multiple genomes of *Lotus*-associated symbionts. By contrast, one *L. japonicus*-associated strain (131-3-5) had nearly the complete set of the type IV protein secretion system (T4SS) genes that were absent in other *L. japonicus*-associated symbionts, except for the *virB7* gene.

Principal component analysis (PCA)

PCA showed genome-wide genetic variations between strains. When based on the presence/absence of orthologs, PC1 and PC2 explained 49.41% and 14.46%, respectively (Supporting Information Fig. S1A), When based on copy number variations, PC1 and PC2 explained 38.05% and 25.16%, respectively (Supporting Information Fig. S1B). In both plots, there were two closely related pairs of strains (131-2-5 and 131-3-5, L-8-3 and L-8-10) and no clear geographical clusters.

We extracted biallelic SNPs from each gene group (core genome 528,912 bp, symbiosis island: 2497 bp) and then performed PCA separately (Fig. 3A, B). PCA based on the SNPs in the core genome showed similar patterns to the one generated by genome-wide ortholog profiles (Fig. 3A); there were two closely related pairs of strains, which were also observed in the ortholog-based plot. By contrast, PCA using the SNPs of the symbiosis island showed a distinct pattern (Fig. 3B): we again found two pairs of closely related strains, but they were different from those observed in the core genome or genome-wide ortholog profiles (113-3-3 and 113-3-9, 131-2-1 and 131-2-5). Overall, PCA suggested that the pattern of relatedness between strains differed markedly between the core genome and the symbiosis island.

Genome-wide view of polymorphisms and gene genealogies

To understand the genome-wide landscape of polymorphisms, we calculated nucleotide diversity and Tajima's *D* statistic for each gene across the genome (Fig. 3C, D; Supporting Information Table S5). Genes in the core genome showed markedly higher nucleotide diversity at both synonymous and nonsynonymous sites (mean 0.4953 and 0.1029, respectively) compared with those of genes on the symbiosis island (mean 0.0250 and 0.006367, respectively). This pattern is consistent with an observation of the smaller number of genes in a previous study (Bamba *et al.* 2019b), and indicates a signature of horizontal gene transfer (HGT) of the symbiosis island. A steep change in diversity at the borders of the symbiosis island suggests that it would have behaved as a unit of HGT.

Tajima's *D* statistic of the core genome (mean 0.6990) was also markedly higher than that of the symbiosis island (mean –0.5304; Fig. 3D; Supporting

Information Table S5). It is also important to note that Tajima's *D* statistic of the core genome was generally positive, indicating the excess of alleles with intermediate frequencies. This pattern would be expected if there were a clear population structure in the core genome, which was indeed observed in our PCA (Fig. 3A). By contrast, Tajima's *D* statistic of the symbiosis island was generally negative, indicating an excess of rare alleles in this region, and to be expected under a scenario of recent selective sweeps including HGT.

Patterns of phylogenetic relationships were also distinct between the core genome and the symbiosis island (Fig. 3E–G, Supporting Information Table S5). In the core genome, the majority of the gene trees (1555/2163 genes) showed a single *Mesorhizobium* clade, whereas there was no gene tree in which *L. japonicus*- or *Lotus*-associated symbionts formed a clade (Fig. 3E; Supporting Information Table S5). By contrast, in the symbiosis island, 14/76 gene trees showed the *L. japonicus*-associated symbionts clade, 55 genes formed the *Lotus*-associated symbionts clade and 14 genes showed the *Mesorhizobium* clade (Fig. 3F; Supporting Information Table S5). These results suggest that *L. japonicus*-associated symbionts mostly have the core genome of *Mesorhizobium*, but their symbiosis island clearly has a different evolutionary origin, supporting the HGT of the symbiosis island into the diverse genetic background of *Mesorhizobium*. In addition, multiple types of topologies in the symbiosis island indicated a history of recombination within the symbiosis island.

Structural variations in the symbiosis island

An alignment by progressiveMauve analysis (Darling *et al.* 2010) provided the genomic regions corresponding to the symbiosis island of the reference strain, *M. japonicum*

MAFF303099. All of the nine newly assembled genomes harbored the symbiosis island. Eight had the symbiosis island on the chromosome, but in strain 131-3-5, the symbiosis island was identified on the plasmid (Fig. 1).

Alignment by progressiveMauve showed that the synteny of symbiosis island of *L. japonicus*-associated symbionts was mostly conserved. Five conservative synteny blocks were identified in the symbiosis island, and all blocks were found in eight of the nine sequenced strains, except for strain 131-3-5, which lacked the whole type III protein secretion system gene cluster (*nol–hrc*). This was consistent with an ortholog search using SonicParanoid (Consentino and Iwasaki 2018), which revealed the lack of all *nol–hrc* genes (Fig. 2).

The largest synteny block, 21 kb away from the start positions of symbiosis islands (symbiosis islands integrase; *intS*) on *M. japonicum* MAFF303099 (Fig. 4A), contained nitrogen-fixing genes (*nif*, *fix* and *fdx*) and several nodulation genes (*nod*, *noe* and *nol*). We note that this largest block was inverted in the strain L-2-11. Between the *nif–fdx* and *nod–nol* blocks, there was a hypervariable region containing many transposase insertions (around 135 kb in *M. japonicum* MAFF303099).

We found that two pairs of strains harbored extremely similar symbiosis islands, even including the hypervariable region (113-3-3 and 113-3-9, 131-2-1 and 131-2-5; Fig. 4B, C). The sequence identity values between 113-3-3 and 113-3-9 and between 131-2-1 and 131-2-5 were 99.6% and 99.9%, respectively. These two pairs corresponded to the pairs identified in the PCA of SNPs from the symbiosis island (Fig. 3B). Because their core genomes are highly different, these data also provide strong evidence for recent HGT in the whole symbiosis island.

425 **Cross-inoculation experiments**

426 We performed a cross-inoculation experiment using 15 *L. japonicus* accessions and nine
427 rhizobial symbionts, resulting in a total of 135 combinations (Supporting Information
428 Tables S1 and S2). We obtained four phenotypes (SDW, NON, NOA and NOAi) from
429 1189 individuals (5–14 per combination; Fig. 5). All phenotypic traits were significantly
430 correlated with each other (Supporting Information Figs S2 and S3, all Pearson's
431 product-moment correlation $P < 2e^{-16}$). All correlation coefficients were positive (0.246
432 to 0.637), except for that between NON and NOAi (–0.351).

433 In the cross-inoculation experiment, we detected significant effects of host,
434 symbionts and their interactions on all four phenotypes (Table 1: all host, symbiont and
435 interaction effects were significant at $P < 0.001$). For all phenotypes, the effect of host
436 was the largest (partial $\eta^2 = 0.200$–0.268), followed by the host × symbiont interaction
437 effect (partial $\eta^2 = 0.145$–0.168) and the effect of symbionts (partial $\eta^2 = 0.038$–0.081;
438 Table 1), indicating that the host phenotypes were more strongly affected by host ×
439 symbiont interactions than the sole symbiont effect. Furthermore, it is worth noting that
440 the combinations of hosts and symbionts from the same localities did not necessarily
441 show higher phenotypic values than nonnative combinations (Supporting Information
442 Table S6), which was not consistent with the pattern expected from local adaptation.

443 To understand what genes or genomic regions of rhizobial symbionts could be
444 responsible for variations in partner quality, we first quantified the rhizobial quality
445 (Supporting Information Fig. S2) and G × G effects (Supporting Information Fig. S4).
446 There were 1.233–1.413 times greater differences between minimum and maximum
447 values of variations in partner quality.

448 Genes previously reported as being involved in symbiosis would be obvious
449 candidates explaining the rhizobial variations in partner quality. Therefore, we first
450 investigated the correlations between the variations in partner quality (represented by
451 mean phenotypic values) and the presence/absence variations of these symbiosis genes.
452 Analysis of variance using the genes in which we found the presence/absence variations
453 (Fig. 2) revealed that none were significantly correlated with variations in partner
454 quality ($P > 0.05$; Supporting Information Table S7). Next, we investigated the
455 correlation between the variations in partner quality and genetic distances of genes that
456 were reported as candidate genes responsible for rhizobial quality (Klinger *et al.* 2016).
457 Mantel testing using these candidate genes found that none were significantly correlated
458 with rhizobial quality variations (Supporting Information Table S8). These results
459 suggest that known symbiosis-related genes would not be responsible for the rhizobial
460 variations in partner quality detected in the cross-inoculation experiment.
461 We then performed an SNP-based association analysis between rhizobial
462 genomes and rhizobial variations in partner quality. We found that SNPs in the core
463 genome were more strongly associated with rhizobial quality than those in the
464 symbiosis islands (Supporting Information Fig. S5). In the linear model for SDW, NON
465 and NOAi, *P* value distributions were strongly skewed toward small values in the core
466 genome (62.6%, 70.7% and 58.5% of SNP *P* values were < 0.05, respectively;
467 Supporting Information Fig. S5A–D). These skewed *P*-value distributions in the core
468 genome presumably resulted from population structure and strong linkage
469 disequilibrium. On the other hand, *P* value distributions were not strongly skewed in the
470 symbiosis islands. Thus, the *P*-values were < 0.05 in 10.8% of SDW SNPs, 5.3% of

NON SNPs, 11.8% of NOA SNPs and 8.8% of NOAi SNPs; Supporting Information Fig. S5E–H).

Next, we performed Mantel tests to determine whether the genetic distance between rhizobial strains was positively correlated with variations in partner quality. The Mantel test results between the variations of rhizobial quality for SDW and NOAi were significantly correlated with the reciprocal of rhizobial core genome kinships ($P < 0.05$ after Bonferroni corrections; Table 2). By contrast, we did not recognize any correlation between variations in rhizobial quality and the kinships of symbiosis islands. These results suggest that variation in the core genome, rather than in the symbiosis islands, could explain the rhizobial variations in partner quality.

Variations in G × G interactions were not significantly correlated with rhizobial core genomes or symbiosis islands (Supporting Information Table S9). However, it is noteworthy that there were variations in G × G interactions between the pairs that appeared to have almost identical symbiosis islands but highly different core genomes (Fig. 4B, C; 113-3-3 and 113-3-9: 131-2-1 and 131-2-5; Supporting Information Fig. S4). There were no significant correlations for G × G variations between these symbiont pairs, except for NOAi of 113-3-3 and 113-3-9 ($P < 0.05$ after Bonferroni corrections; Supporting Information Table S10). Because the symbiosis islands are nearly identical, the observed G × G variations should due to genetic variation in the core genome.

DISCUSSION

De novo assembled genomes of *Lotus japonicus*-associated nodule bacteria

While there have been several attempts to sequence the whole genomes of nodule bacteria (Kaneko *et al.* 2000, 2002; Amadou *et al.* 2008; Lee *et al.* 2008; Reeve *et al.* 2010ab, 2015; Ramsay *et al.* 2013; Moulin *et al.* 2014; Wang *et al.* 2014; Shimoda *et al.* 2016; Nagymihály *et al.* 2017; Liang *et al.* 2018), *de novo* sequencing of multiple strains associated with a single plant species has been uncommon. Here, we performed a whole-genome sequencing analysis of nine *L. japonicus*-associated symbionts by exploiting both long-read (Oxford Nanopore MinION) and short-read (Illumina HiSeq) sequencers, which enabled us to generate high quality *de novo* assembled genomes for each strain.

Comparative genomic analyses of these sequenced genomes provided clear evidence for HGT of the symbiosis island. First, patterns of phylogenetic relationships were distinct between the core genome and the symbiosis island (Fig. 3E, F; Supporting Information Table S5). In the core genome, there was no gene forming an *L. japonicus*- or *Lotus*-associated symbiont clade, but in the symbiosis island, 14 and 55 of 76 genes showed *L. japonicus*- or *Lotus*-associated symbiont clades, respectively, indicating that the evolutionary origin of the symbiosis island was clearly different from that of the core genome (Fig. 3E, Supporting Information Table S5). Second, we observed a marked decline in the nucleotide diversity π statistic at synonymous and nonsynonymous sites and negative Tajima's *D* statistic, which are also evidence for HGT (Fig. 3C, D). A steep change in diversity at the borders of the symbiosis island suggests that it would have behaved as a unit of HGT, while the multiple topologies of

513 gene trees in the symbiosis island also indicate an evolutionary history of internal
514 recombination events. Third, there was a pair of strains harboring almost identical
515 sequences of the symbiosis island, but with highly distinct core genome backgrounds,
516 which is also a signature of recent HGT (Fig. 4B, C). Including our previous study of *L.*
517 *japonicus*-associated nodule bacteria (Bamba *et al.* 2019b), there have been several
518 studies demonstrating HGT of symbiosis islands based on sequences of a few genes
519 (Barcellos *et al.* 2007; Steenkamp *et al.* 2008; Menna and Hungria 2011; Koppell and
520 Parker 2012; Parker and Rousteau 2014; Lemaire *et al.* 2015; Bamba *et al.* 2016). Thus,
521 our *de novo* assembled genome data have provided multiple signatures of HGT at an
522 unprecedented level of resolution.
523 The assembled genomes also revealed blocks of conserved synteny as well as
524 extensive structural rearrangements in the symbiosis island. While Shimoda *et al.*
525 (2016) showed that there were three conserved regions (*nif*, *nod* and type III protein
526 secretion system) in the symbiosis islands of *Lotus*-associated symbionts, we found five
527 conserved blocks and many rearrangements, including an inversion and loss of genes
528 and gene clusters. As these structural rearrangements were found in the symbiosis island
529 of bacteria associated with a single legume species, rearrangements or gene gain/loss
530 could occur over a short time. We also found evidence of recombination within the
531 symbiosis island based on the topologies of gene trees (Fig. 3E, F). Signatures of gene
532 gain/loss and recombination have also been reported for several rhizobial genera
533 (*Burkholderia*, De Meyer *et al.* 2016; *Bradyrhizobium*, Sugawara *et al.* 2013; Bouznif *et*
534 *al.* 2019; Porter *et al*. 2019), and the genomic insight of the symbiosis island of *L.*
535 *japonicus*-associated symbionts is consistent with the emerging perspective that

recombination and gene gain/loss are not rare events in rhizobial symbiosis islands/plasmids (Porter *et al*. 2019).

Genomic regions associated with variations in partner quality

By integrating the data of cross-inoculation experiments and assembled whole genomes, we investigated which genomic regions could be responsible for variations in partner quality. We found that plant growth was significantly influenced by host genotypes, symbiont genotypes and host × symbiont (G × G) interactions (Table 1). We then examined whether rhizobial variations in partner quality were explained by the following genetic factors: (i) presence/absence variation of symbiosis genes, (ii) SNPs in the rhizobial genomes and (iii) genetic kinships of the core genome and the symbiosis island. We found that the rhizobial core genome variations explained the rhizobial variations in partner quality: Mantel tests showed that the quality variations in rhizobial symbionts and their core genome kinship were significantly correlated in two phenotypes (SDW and NOAi; Spearman rank correlations $P < 0.05$ after Bonferroni correction; Table 2), whereas the kinship of the symbiosis island was not. In the SNP-based association analyses, *P*-value distributions of core genomes SNPs were strongly skewed toward small values (Supporting Information Fig. S5), suggesting that many SNPs across the genomes are correlated with variations in partner quality, consistent with the results of Mantel tests. Such strongly skewed *P*-value distributions possibly arise from extensive genome-wide linkage disequilibrium. Bacterial genomes generally show strong linkage disequilibrium given their asexual reproduction (Chen *et al.* 2015), which could make it difficult to pinpoint the responsible regions/genes on the genome using association-based analysis.

Our finding of significant correlations between core genome variations and the rhizobial variations in partner quality might partly explain why such variations persist in legume–rhizobia mutualisms. In *L. japonicus*-associated symbionts, massive genetic variations in the core genome would be maintained by recurrent HGT of the symbiosis islands into diverse *Mesorhizobium* bacterial strains (Bamba *et al.* 2019b). Local *Mesorhizobium* communities could thus serve as a source of standing genetic variation of core genomes, which might prevent variations in partner quality from fixing even under the presence of selection favoring high-quality partners: i.e. a stabilizing mechanism in mutualisms (Heath and Stinchcombe 2014). In the context of the mutation–selection balance model (Van Dyken *et al.* 2011; Smith *et al.* 2014), our study serves to illuminate the role of HGT among symbiosis islands that indirectly supply mutations in core genomes contributing to variations in partner quality.

A few studies have reported that variations in the symbiosis islands explain rhizobial variations in partner quality, unlike our finding in *L. japonicus*-associated bacteria (Klinger *et al.* 2016; Porter *et al.* 2019). Klinger *et al.* (2016) showed that variations in partner quality are explained by variations in the *nifH*, *nifA*, *fixC*, *nodB* and Rleg_4928 (*fixB*) genes. Although all these genes are present in the genomes of *L. japonicus*-associated symbionts (Fig. 2), genetic distances of these genes among strains and variations in rhizobial quality were not significantly correlated, suggesting that these genes do not explain the variations in partner quality in our system (Supporting Information Table S8). Porter *et al.* (2019) showed that the absence of rhizobial symbiosis genes lessens the quality of rhizobial symbionts. However, in our nine sequenced rhizobial genomes, none of the symbiotic genes showing the presence/absence of variations—all type III protein secretion system-related genes (Fig.

584 2)—were significantly correlated with rhizobial quality variations ($P < 0.05$ after
585 Bonferroni correction; Supporting Information Table S7). We note that type III protein
586 secretion system-related genes might not be involved in interactions between *L.*
587 *japonicus* and *Meshorhizobium* because dysfunctional mutants of such genes in *M.*
588 *japonicum* did not show phenotypic changes in terms of nodule-forming ability for *L.*
589 *japonicus* B129 (Okazaki *et al.* 2010), although their effects have been detected in other
590 *Lotus* species (Okazaki *et al.* 2010; Mercante *et al.* 2015).

591 We speculate that the contrasting findings between our study and previous ones
592 might have arisen in part from differences in sampling schemes and in the unique
593 history of *L. japonicus*-associated symbionts. Both Klinger *et al.* (2016) and Porter *et al.*
594 (2019) focused on rhizobial populations harboring similar core genomes. Porter *et al.*
595 (2019) used 38 strains possessing a similar core genome as a recombining population,
596 and Klinger *et al.* (2016) analyzed strains collected from nitrogen-enriched
597 experimental fields, and the nucleotide diversity of their core genomes was as low as
598 that of the symbiosis islands. By contrast, data compilation by Bamba *et al.* (2019b)
599 revealed that one of the notable characteristics of the *L. japonicus*-associated rhizobia in
600 Japan is the presence of highly diverse core genomes and the extremely low nucleotide
601 diversity of the symbiosis island. This possibly reflects recent and recurrent HGT of the
602 symbiosis island associated with the population expansion of *L. japonicus* into the Japan
603 archipelago over several thousand years (Bamba *et al.* 2019b). Therefore, it is possible
604 that the symbiosis islands of *L. japonicus*-associated bacteria analyzed in this study
605 have experienced genetic bottlenecks when *L. japonicus* migrated into the Japan
606 archipelago and are too homogeneous to serve as a source for variations in partner
607 quality.

608

609 **G × G interactions and local adaptation**

610 In our cross-inoculation experiment, we found that plant growth was significantly
611 influenced by host × symbiont (G × G) interactions, even more than symbiont genotypes
612 (Table 1), as is also observed in a few other symbiosis systems (Heath and Tiffin 2007;
613 Barrett *et al.* 2016). Such G × G interactions are suggested to underlie the selective
614 explanations for the persistence of variations in partner quality in mutualisms (Heath
615 and Stinchcombe, 2014), and our cross-inoculation experimental results support this
616 scenario.

617 We found complex G × G interactions even between two rhizobial pairs sharing
618 almost identical symbiosis islands (113-3-3 vs. 113-3-9; 131-2-1 vs. 131-2-5; Fig. 4B, C
619 and Supporting Information Fig. S4), strongly suggesting that such variations have
620 arisen in part from variations in rhizobial core genomes. Previous studies on G × G
621 interactions in legume–rhizobia mutualisms used relatively few genes, so it remained
622 unclear which genes/genomic regions were responsible for such interactions (Heath and
623 Tiffin 2007; Barret *et al.* 2015;). Here, we provide clear evidence supporting the
624 contribution of rhizobial core genome variations. However, we did not observe
625 statistically significant correlations between the variations in G × G interactions and
626 rhizobial genomic variations based on the Mantel test (Supporting Information Table
627 S9). While G × G interactions should have a genetic basis (Table 1), they might be
628 governed by polygenic factors that would not be detectable at our relatively small
629 experimental scale (15 rhizobia × 9 legume combinations).

630 Such G × G interactions in a spatial context have been a hotly debated issue in
631 legume–rhizobia mutualisms (Heath and Tiffin 2007; Heath 2010; Porter *et al.* 2011;

632 Heath *et al.* 2012; Ehinger *et al.* 2014; Harrison *et al.* 2017). The geographic mosaic
633 theory of coevolution states that the outcome of reciprocal selection between a
634 particular genotype of one species and a genotype of an interacting species will differ
635 among ecologically distinct locations (Thompson 1994, 1997, 2005; Forde *et al.* 2004;
636 Decaestecker *et al.* 2007; Laine *et al.* 2014;). According to this theory, if legume–
637 rhizobia mutualistic interactions coevolve locally, native rhizobial genotypes are
638 expected to outperform foreign rhizobial genotypes when associated with host
639 genotypes originating from the same locations. In our experiments, the NOA and SDW
640 data can be considered fitness proxies (Ratcliff *et al.* 2012; Younginger *et al.* 2017), but
641 the combinations of hosts and symbionts from the same localities did not show higher
642 phenotypic values in either of them (Supporting Information Table S6), which is not
643 consistent with the pattern expected from local adaptation. As is also discussed for the
644 maintenance of variations in partner quality, recurrent HGT of the symbiosis islands
645 into diverse *Mesorhizobium* core genomes might explain in part the absence of local
646 adaptations in the *L. japonicus* associated-symbionts, given that core genome variations
647 are strongly correlated with rhizobial variations in partner quality. With local and
648 recurrent HGT of symbiosis islands, core genome variations underlying variations in
649 partner quality might be prevented from fixation even under the presence of local
650 reciprocal selection between plants and rhizobia. By integrating the full-genome
651 sequencing of rhizobial strains and cross-inoculation experiments, this study has
652 demonstrated a scenario of how variations in partner quality could be maintained in the
653 presence of selection and HGT of symbiosis islands. More genetic studies from a plant
654 perspective would be valuable for the understanding of coevolution between plants and
655 rhizobia. This has now become possible for *L. japonicus*, where full-genome

656 resequenced data have become available for hundreds of natural accessions (Shah *et al.*

657 2020).

ACKNOWLEDGMENTS

Wild accessions of *L. japonicus* used in this research were provided by the National BioResource Project "Lotus/Glycine" of the Ministry of Education, Culture, Sports, Science and Technology, Japan.

FUNDING

This work was supported by a Research Fellowship from the Japan Society for the Promotion of Science for Young Scientists [17J04284 to MB], JSPS KAKENHI [grant number 19H03271 to TT]; Ministry of Education, Culture, Sports, Science and Technology (KAKENHI) [grant numbers 17H05833, 18H04813, and 19H04851 to TT]; the Sumitomo Foundation research grant [161380 to TT].

670 **REFERENCES**

671 Amadou C, Mangenot S, Glew M *et al.* Genome sequence of the β-rhizobium
672 *Cupriavidus taiwanensis* and comparative genomics of rhizobia. *Genome Res*
673 2008;18:1472–83.

674 Archetti M, Scheuring I, Hoffman M *et al.* Economic game theory for mutualism and
675 cooperation. *Ecol Lett* 2011;14:1300–12.

676 Aziz RK, Bartels D, Best A *et al.* The RAST Server: Rapid annotations using
677 subsystems technology. *BMC Genomics* 2008;9:1–15.

678 Bamba M, Nakata S, Aoki S *et al.* Wide distribution range of rhizobial symbionts
679 associated with pantropical sea-dispersed legumes. *Antonie Van Leeuwenhoek*
680 2016;109:1605–14.

681 Bamba M, Kawaguchi YW, Tsuchimatsu T. Plant adaptation and speciation studied by
682 population genomic approaches. *Dev Growth, Differ* 2019a;61:12–24.

683 Bamba M, Aoki S, Kajita T *et al.* Exploring Genetic Diversity and Signatures of
684 Horizontal Gene Transfer in Nodule Bacteria Associated with *Lotus japonicus* in
685 Natural Environments. *Mol Plant-Microbe Interact* 2019b;32:1110–20.

686 Bankevich A, Nurk S, Antipov D *et al.* SPAdes: A new genome assembly algorithm and
687 its applications to single-cell sequencing. *J Comput Biol* 2012;19:455–77.

688 Barcellos FG, Menna P, Batista JSDS *et al.* Evidence of horizontal transfer of symbiotic
689 genes from a *Bradyrhizobium japonicum* inoculant strain to indigenous diazotrophs

690 *Sinorhizobium* (*Ensifer*) *fredii* and *Bradyrhizobium elkanii* in a Brazilian savannah
691 soil. *Appl Environ Microbiol* 2007;73:2635–43.

692 Barrett LG, Bever JD, Bissett A *et al.* Partner diversity and identity impacts on plant
693 productivity in *Acacia*-rhizobial interactions. *J Ecol* 2015;103, DOI:
694 10.1111/1365-2745.12336.

695 Barrett LG, Zee PC, Bever JD *et al.* Evolutionary history shapes patterns of mutualistic
696 benefit in *Acacia*-rhizobial interactions. *Evolution* 2016;70:1473–85.

697 Bouznif B, Guefrachi I, Rodríguez de la Vega RC *et al.* Phylogeography of the
698 *Bradyrhizobium* spp. Associated With Peanut, *Arachis hypogaea*: Fellow Travelers
699 or New Associations? *Front Microbiol* 2019;10:1–13.

700 Brettin T, Davis JJ, Disz T *et al.* RASTtk: A modular and extensible implementation of
701 the RAST algorithm for building custom annotation pipelines and annotating
702 batches of genomes. *Sci Rep* 2015;5, DOI: 10.1038/srep08365.

703 Broughton WJ, Dilworth MJ. Control of leghaemoglobin synthesis in snake beans.
704 *Biochem J* 1971;125:1075–80.

705 Bull JJ, Rice WR. Distinguishing mechanisms for the evolution of co-operation. *J Theor*
706 *Biol* 1991;149:63–74.

707 Chen PE, Shapiro BJ. The advent of genome-wide association studies for bacteria. *Curr*
708 *Opin Microbiol* 2015;25:17–24.

709 Cosentino S, Iwasaki W. SonicParanoid: Fast, accurate and easy orthology inference.
710 *Bioinformatics* 2019;35:149–51.

711 Cox MP, Peterson DA, Biggs PJ. SolexaQA: At-a-glance quality assessment of Illumina
712 second-generation sequencing data. *BMC Bioinformatics* 2010;11:485.

713 Darling AE, Mau B, Perna NT. Progressivemauve: Multiple genome alignment with
714 gene gain, loss and rearrangement. *PLoS One* 2010;5, DOI:
715 10.1371/journal.pone.0011147.

716 Darriba D, Posada D, Kozlov AM *et al.* ModelTest-NG: A New and Scalable Tool for
717 the Selection of DNA and Protein Evolutionary Models. *Mol Biol Evol* 2019, DOI:
718 10.1093/molbev/msz189.

719 De Coster W, D'Hert S, Schultz DT *et al.* NanoPack: Visualizing and processing long-
720 read sequencing data. *Bioinformatics* 2018;34:2666–9.

721 De Meyer SE, Briscoe L, Martínez-Hidalgo P *et al.* Symbiotic *Burkholderia* species
722 show diverse arrangements of *nif*/*fix* and nod genes and lack typical High-Affinity
723 cytochrome *cbb3* Oxidase genes. *Mol Plant-Microbe Interact* 2016;29:609–19.

724 Decaestecker E, Gaba S, Raeymaekers JAM *et al.* Host-parasite "Red Queen" dynamics
725 archived in pond sediment. *Nature* 2007;450:870–3.

726 Denison RF, Toby Kiers E. Why are most rhizobia beneficial to their plant hosts, rather
727 than parasitic? *Microbes Infect* 2004;6:1235–9.

728 Ehinger M, Mohr TJ, Starcevich JB *et al*. Specialization-generalization trade-off in a
729 *Bradyrhizobium* symbiosis with wild legume hosts. *BMC Ecol* 2014;14:8.

730 Forde SE, Thompson JN, Bohannan BJM. Adaptation varies through space and time in
731 a coevolving host-parasitoid interaction. *Nature* 2004;431:841–4.

732 Frederickson ME. Rethinking mutualism stability: cheaters and the evolution of
733 sanctions. *Q Rev Biol* 2013;88:269–95.

734 Ghoul M, Griffin AS, West SA. Toward an evolutionary definition of cheating.
735 *Evolution (N Y)* 2014;68:318–31.

736 Handberg K, Stougaard J. *Lotus japonicus* , an autogamous, diploid legume species for
737 classical and molecular genetics. *Plant J* 1992;2:487–96.

738 Harrison TL, Wood CW, Heath KD *et al*. Geographically structured genetic variation in
739 the *Medicago lupulina–Ensifer* mutualism. *Evolution (N Y)* 2017;71:1787–801.

740 Heath KD. Intergenomic epistasis and coevolutionary constraint in plants and rhizobia.
741 *Evolution (N Y)* 2010;64:1446–58.

742 Heath KD, Burke P V., Stinchcombe JR. Coevolutionary genetic variation in the
743 legume-rhizobium transcriptome. *Mol Ecol* 2012;21:4735–47.

744 Heath KD, Stinchcombe JR. Explaining mutualism variation: a new evolutionary
745 paradox? *Evolution* 2014;68:309–17.

746 Heath KD, Tiffin P. Context dependence in the coevolution of plant and rhizobial
747 mutualists. *Proc R Soc B* 2007;274:1905–12.

748 Hubber A, Vergunst AC, Sullivan JT *et al.* Symbiotic phenotypes and translocated
749 effector proteins of the *Mesorhizobium loti* strain R7A VirB/D4 type IV secretion
750 system. *Mol Microbiol* 2004;54:561–74.

751 Kaneko T, Nakamura Y, Sato S *et al.* Complete genome structure of the nitrogen-fixing
752 symbiotic bacterium *Mesorhizobium loti*. *DNA Res* 2000;7:381–406.

753 Kaneko T, Nakamura Y, Sato S *et al.* Complete genomic sequence of nitrogen-fixing
754 symbiotic bacterium *Bradyrhizobium japonicum* USDA110. *DNA Res* 2002;9:189–
755 97.

756 Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7:
757 Improvements in Performance and Usability. *Mol Biol Evol* 2013;30:772–80.

758 Kiers ET, Rousseau RA, West SA *et al.* Host sanctions and the legume–rhizobium
759 mutualism. *Nature* 2003;425:78–81.

760 Klinger CR, Lau JA, Heath KD. Ecological genomics of mutualism decline in nitrogen-
761 fixing bacteria. *Proc R Soc B* 2015;283.

762 Koppell JH, Parker M a. Phylogenetic clustering of *Bradyrhizobium* symbionts on
763 legumes indigenous to North America. *Microbiology* 2012;158:2050–9.

764 Kouchi H, Shimomura K, Hata S *et al.* Large-scale analysis of gene expression profiles
765 during early stages of root nodule formation in a model legume, *Lotus japonicus*.
766 *DNA Res* 2004;11:263–74.

767 Kozlov AM, Darriba D, Flouri T *et al.* RAxML-NG: a fast, scalable and user-friendly
768 tool for maximum likelihood phylogenetic inference. *Bioinformatics* 2019:1–5.

769 Kumar S, Stecher G, Peterson D *et al.* MEGA-CC: Computing core of molecular
770 evolutionary genetics analysis program for automated and iterative data analysis.
771 *Bioinformatics* 2012;28:2685–6.

772 Laine AL, Burdon JJ, Nemri A *et al.* Host ecotype generates evolutionary and
773 epidemiological divergence across a pathogen metapopulation. *Proc R Soc B Biol*
774 *Sci* 2014;281, DOI: 10.1098/rspb.2014.0522.

775 Lee KB, De Backer P, Aono T *et al.* The genome of the versatile nitrogen fixer
776 *Azorhizobium caulinodans* ORS571. *BMC Genomics* 2008;9:1–14.

777 Lemaire B, Van Cauwenberghe J, Chimphango S *et al.* Recombination and horizontal
778 transfer of nodulation and ACC deaminase (*acdS*) genes within Alpha- and
779 Betaproteobacteria nodulating legumes of the Cape Fynbos biome. *FEMS*
780 *Microbiol Ecol* 2015;91:1–11.

781 Leonard LT. A simple assembly for use in the testing of cultures of rhizobia. J Bacteriol
782 1943;45:523.

783 Liang J, Hoffrichter A, Brachmann A *et al.* Complete genome of *Rhizobium*
784 *leguminosarum* Norway, an ineffective *Lotus* micro-symbiont. *Stand Genomic Sci*
785 2018;13:1–11.

786 Lippert C, Casale FP, Rakitsch B *et al.* LIMIX: genetic analysis of multiple traits.
787 *bioRxiv* 2014:003905.

788 Madsen LH, Tirichine L, Jurkiewicz A *et al.* The molecular network governing nodule
789 organogenesis and infection in the model legume *Lotus japonicus*. *Nat Commun*
790 2010;1, DOI: 10.1038/ncomms1009.

791 Maekawa T, Maekawa-Yoshikawa M, Takeda N *et al.* Gibberellin controls the
792 nodulation signaling pathway in *Lotus japonicus*. *Plant J* 2009;58:183–94.

793 Matsui M, Iwasaki W. Graph Splitting: A Graph-Based Approach for Superfamily-
794 Scale Phylogenetic Tree Reconstruction. *Syst Biol* 2019;0:1–15.

795 Menna P, Hungria M. Phylogeny of nodulation and nitrogen-fixation genes in
796 *Bradyrhizobium*: Supporting evidence for the theory of monophyletic origin, and
797 spread and maintenance by both horizontal and vertical transfer. *Int J Syst Evol*
798 *Microbiol* 2011;61:3052–67.

799 Mercante V, Duarte CM, Sánchez CM *et al.* The absence of protein Y4yS affects
800 negatively the abundance of T3SS *Mesorhizobium loti* secretin, RhcC2, in bacterial
801 membranes. *Front Plant Sci* 2015;6:1–12.

802 Moulin L, Klonowska A, Caroline B *et al.* Complete Genome sequence of *Burkholderia*
803 *phymatum* STM815T, a broad host range and efficient nitrogen-fixing symbiont of
804 *Mimosa* species. *Stand Genomic Sci* 2015;9:763–74.

805 Nagymihály M, Vásarhelyi BM, Barrière Q *et al.* The complete genome sequence of
806 *Ensifer meliloti* strain CCMM B554 (FSM-MA), a highly effective nitrogen-fixing
807 microsymbiont of *Medicago truncatula* Gaertn. *Stand Genomic Sci* 2017;12:1–9.

808 Nishida H, Handa Y, Tanaka S *et al.* Expression of the *CLE-RS3* gene suppresses root
809 nodulation in *Lotus japonicus. J Plant Res* 2016;129:909–19.

810 Nishida H, Tanaka S, Handa Y *et al.* A NIN-LIKE PROTEIN mediates nitrate-induced
811 control of root nodule symbiosis in *Lotus japonicus*. *Nat Commun* 2018;9:1–14.

812 Okazaki S, Okabe S, Higashi M *et al.* Identification and functional analysis of type III
813 effector proteins in *Mesorhizobium loti*. *Mol Plant-Microbe Interact* 2010;23:223–
814 34.

815 Overbeek R, Olson R, Pusch GD *et al.* The SEED and the Rapid Annotation of
816 microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res*
817 2014;42:206–14.

818 Parker M a., Rousteau A. Mosaic origins of *Bradyrhizobium* legume symbionts on the
819 Caribbean island of Guadeloupe. *Mol Phylogenet Evol* 2014;77:110–5.

820 Porter SS, Faber-Hammond J, Montoya AP *et al.* Dynamic genomic architecture of
821 mutualistic cooperation in a wild population of *Mesorhizobium*. *ISME J*
822 2019;13:301–15.

823 Porter SS, Stanton ML, Rice KJ. Mutualism and adaptive divergence: Co-invasion of a
824 heterogeneous grassland by an exotic legume-rhizobium symbiosis. *PLoS One*
825 2011;6, DOI: 10.1371/journal.pone.0027935.

826 R Core Team R: A language and environment for statistical computing. 2019.

827 Raffaele S, Farrer RA, Cano LM *et al.* Genome evolution following host jumps in the
828 irish potato famine pathogen lineage. *Science (80-)* 2010;330:1540–3.

829 Ramsay JP, Major AS, Komarovsky VM *et al.* A widely conserved molecular switch
830 controls quorum sensing and symbiosis island transfer in *Mesorhizobium loti*
831 through expression of a novel antiactivator. *Mol Microbiol* 2013;87:1–13.

832 Ratcliff WC, Underbakke K, Denison RF. Measuring the fitness of symbiotic rhizobia.
833 *Symbiosis* 2011;55:85–90.

834 Reeve W, Chain P, O'Hara G *et al.* Complete genome sequence of the *Medicago*
835 microsymbiont *Ensifer* (*Sinorhizobium*) *medicae* strain WSM419. *Stand Genomic*
836 *Sci* 2010a;2:77–86.

837 Reeve W, O'Hara G, Chain P *et al.* Complete genome sequence of *Rhizobium*
838 *leguminosarum* bv *trifolii* strain WSM2304, an effective microsymbiont of the

South American clover *Trifolium polymorphum*. *Stand Genomic Sci* 2010b;2:66–76.

Reeve W, Sullivan J, Ronson C *et al*. High-Quality draft genome sequence of the *Lotus* spp. microsymbiont *Mesorhizobium loti* strain CJ3Sym. *Stand Genomic Sci* 2015;10:54.

Sachs JL, Ehinger MO, Simms EL. Origins of cheating and loss of symbiosis in wild *Bradyrhizobium*. *J Evol Biol* 2010;23:1075–89.

Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 2011;27:863–4.

Segura V, Vilhjálmsson BJ, Platt A *et al.* An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nat Genet* 2012;44:825–30.

Shah N, Wakabayashi T, Kawamura Y *et al*. Extreme genetic signatures of local adaptation during *Lotus japonicus* colonization. *Nat Commun* 2020;**11**, DOI: 10.1038/s41467-019-14213-y |.

Shimoda Y, Hirakawa H, Sato S *et al*. Whole-Genome Sequence of the Nitrogen-Fixing Symbiotic Rhizobium *Mesorhizobium loti* Strain TONO. *Genome Announc* 2016;4:e01016-16.

857 Simonsen AK, Stinchcombe JR. Standing genetic variation in host preference for
858 mutualist microbial symbionts. *Proc R Soc B Biol Sci* 2014;281, DOI:
859 10.1098/rspb.2014.2036.

860 Smith J, Van Dyken JD, Velicer GJ. Nonadaptive processes can create the appearance
861 of facultative cheating in microbes. *Evolution (N Y)* 2014;68:816–26.

862 Souza RC, Del Rosario Quispe Saji G, Costa MO *et al.* AtlasT4SS: A curated database
863 for type IV secretion systems. *BMC Microbiol* 2012;12, DOI: 10.1186/1471-2180-
864 12-172.

865 Soyano T, Kouchi H, Hirota A *et al.* NODULE INCEPTION Directly Targets *NF-Y*
866 Subunit Genes to Regulate Essential Processes of Root Nodule Development in
867 *Lotus japonicus*. *PLoS Genet* 2013;9, DOI: 10.1371/journal.pgen.1003352.

868 Steenkamp ET, Stepkowski T, Przymusiak A *et al.* Cowpea and peanut in southern
869 Africa are nodulated by diverse *Bradyrhizobium* strains harboring nodulation genes
870 that belong to the large pantropical clade common in Africa. *Mol Phylogenet Evol*
871 2008;48:1131–44.

872 Steidinger BS, Bever JD. The coexistence of hosts with different abilities to
873 discriminate against cheater partners: An evolutionary game-theory approach. *Am*
874 *Nat* 2014;183:762–70.

875 Steinegger M, Söding J. MMseqs2 enables sensitive protein sequence searching for the
876 analysis of massive data sets. *Nat Biotechnol* 2017;35:1026–8.

877 Sugawara M, Epstein B, Badgley BD *et al.* Comparative genomics of the core and
878 accessory genomes of 48 *Sinorhizobium* strains comprising five genospecies.
879 *Genome Biol* 2013;14:R17.

880 Suzuki A, Suriyagoda L, Shigeyama T *et al. Lotus japonicus* nodulation is
881 photomorphogenetically controlled by sensing the red/far red (R/FR) ratio through
882 jasmonic acid (JA) signaling. *Proc Natl Acad Sci* 2011;108:16837–42.

883 Szczyglowski K, Shaw RS, Wopereis J *et al.* Nodule organogenesis and symbiotic
884 mutants of the model legume *Lotus japonicus*. *Mol Plant-Microbe Interact*
885 1998;11:684–97.

886 Tamura K, Battistuzzi FU, Billing-Ross P *et al.* Estimating divergence times in large
887 molecular phylogenies. *Proc Natl Acad Sci U S A* 2012;109:19333–8.

888 Thompson JN. Evaluating the dynamics of coevolution among geographically
889 structured populations. *Ecology* 1997;78:1619–23.

890 Thompson JN. *The Coevolutionary Process*. University of Chicago Press, 1994.

891 Thompson JN. Coevolution: The Geographic Mosaic of Coevolutionary Arms Races.
892 *2Current Biol* 2005;15:R9920-R994.

893 Thrall PH, Burdon JJ, Woods MJ. Variation in the effectiveness of symbiotic
894 associations between native rhizobia and temperate Australian legumes:
895 Interactions within and between genera. *J Appl Ecol* 2000;37:52–65.

896 Tsushima A, Gan P, Kumakura N *et al.* Genomic plasticity mediated by transposable
897 elements in the plant pathogenic fungus *Colletotrichum higginsianum*. *Genome*
898 *Biol Evol* 2019;11:1487–500.

899 Van Dyken JD, Linksvayer TA, Wade MJ. Kin selection-mutation balance: A model for
900 the origin, maintenance, and consequences of social cheating. *Am Nat*
901 2011;177:288–300.

902 Virtanen P, Gommers R, Oliphant TE *et al.* SciPy 1.0--Fundamental Algorithms for
903 Scientific Computing in Python. 2019:1–22.

904 Walker BJ, Abeel T, Shea T *et al.* Pilon: An integrated tool for comprehensive
905 microbial variant detection and genome assembly improvement. *PLoS One* 2014;9,
906 DOI: 10.1371/journal.pone.0112963.

907 Wang S, Hao B, Li J *et al.* Whole-genome sequencing of *Mesorhizobium huakuii*
908 7653R provides molecular insights into host specificity and symbiosis island
909 dynamics. *BMC Genomics* 2014;15:1–17.

910 Waterhouse RM, Seppey M, Simao FA *et al.* BUSCO applications from quality
911 assessments to gene prediction and phylogenomics. *Mol Biol Evol* 2018;35:543–8.

912 Wick RR, Judd LM, Gorrie CL *et al.* Completing bacterial genome assemblies with
913 multiplex MinION sequencing. *Microb Genomics* 2017:1–7.

914 Wick RR, Schultz MB, Zobel J *et al.* Bandage: Interactive visualization of de novo
915 genome assemblies. *Bioinformatics* 2015;31:3350–2.

916 Younginger BS, Sirová D, Cruzan MB *et al*. Is Biomass a Reliable Estimate of Plant

917 Fitness? *Appl Plant Sci* 2017;5:1600094.

918 **Table 1.** ANOVAs for growth and nodulation of plants in the inoculum treatments.

	d.f.	SDW[a]	NON[b]	NOA[c]	NOAi[d]
Host	14	0.244***	0.228***	0.268***	0.200***
Inoculant	8	0.093***	0.048***	0.038***	0.081***
Host x Inoculant	112	0.161***	0.168***	0.145***	0.162***

[a] Shoot dry weight, [b] Number of nodules, [c] Total size of nodules, and [d] Nodule size per one nodule.

Numerators indicate degree of freedom for d.f and partial η^2 for each phenotype

An astarisk (***) indicates significance at $P < 0.001$

Table 2. Mantel test between partner quality variations and kinships of rhizobial genomes

	Core[a]		Sym[b]	
	Coefficient[a]	P-value[b]	Coefficient[a]	P-value[b]
SDW	0.6561	0.0240	0.1788	1.0000
NON	0.6556	0.0560	-0.0821	1.0000
NOA	0.1030	1.0000	0.0729	1.0000
NOAi	0.9333	0.0008	0.0361	1.0000

[a]Numerators indicate Spearman's rank correlation coefficient.

[b]Numerators indicate P-value after the Bonferroni correction

919

FIGURES

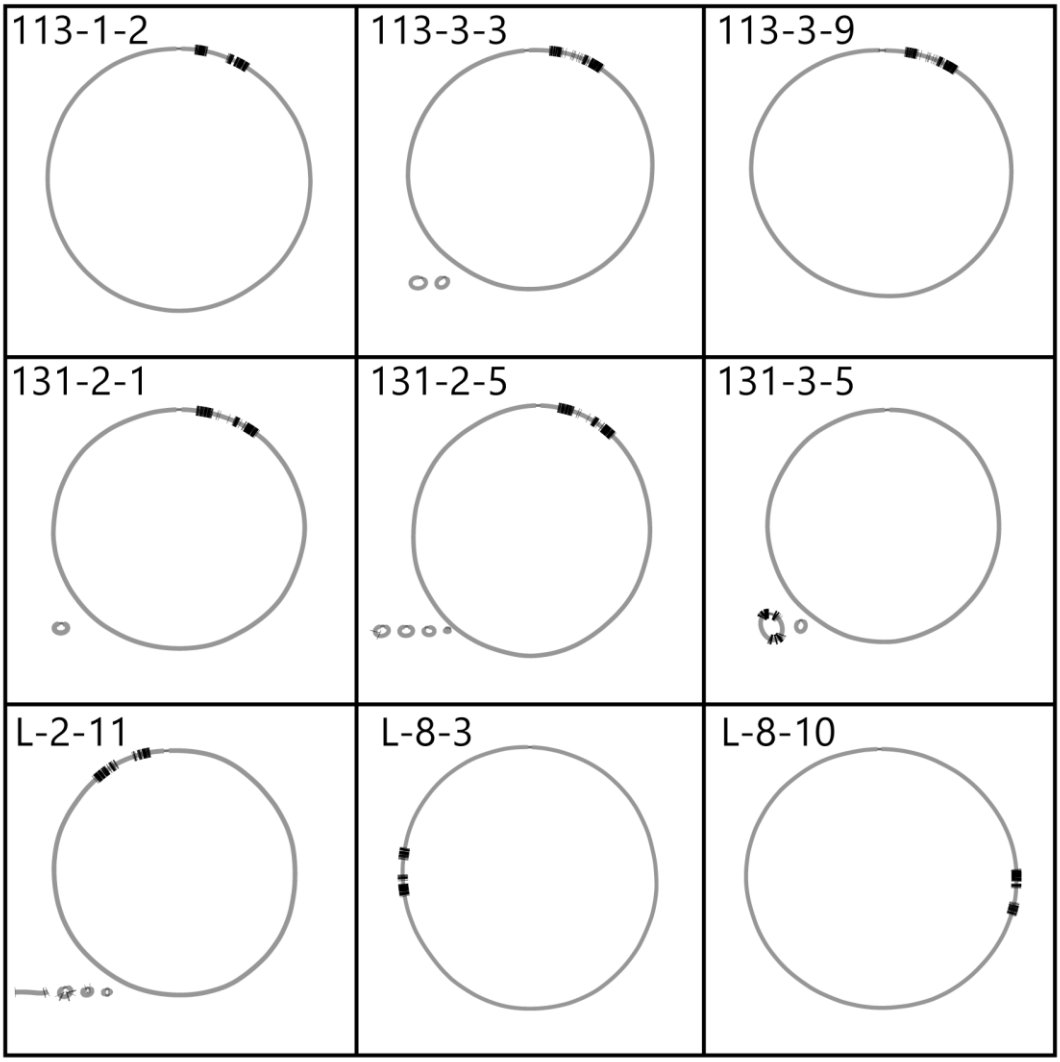

Figure 1. Assembly graphs of nine newly assembled rhizobial genomes. Gray circles and bars indicate assembled genomes. Thick black bars on the genomes indicate genes identified as homologs of genes on the symbiosis island of the reference strain, *M. japonicum* MAFF303099.

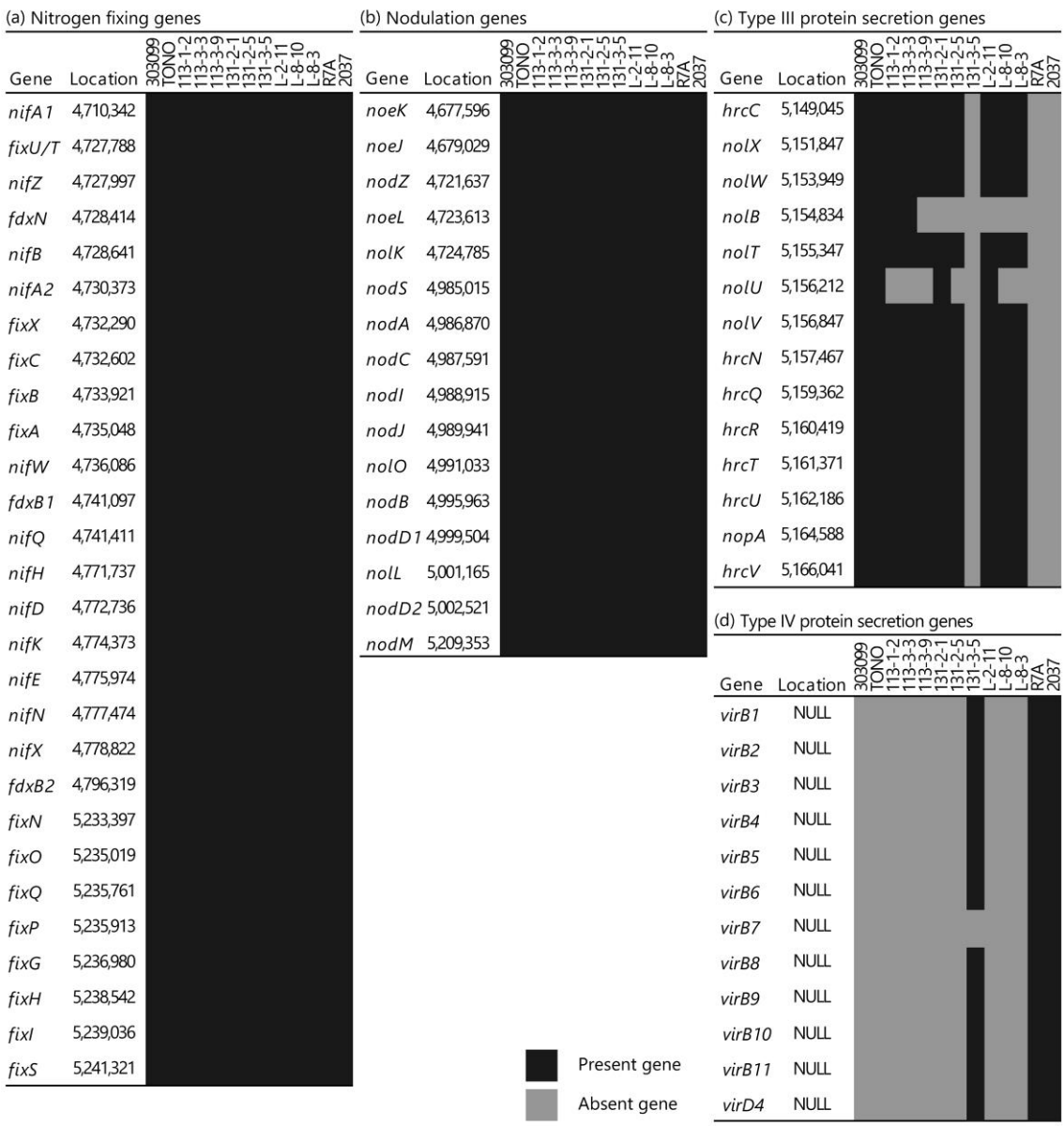

Figure 2. Presence/absence of variations among genes reported to be related to the symbiosis: (a) nitrogen-fixing genes; (b) nodulation genes; (c) type III protein secretion system genes; and (d) type IV protein secretion system genes. Each row indicates rhizobial strains with abbreviations as follows: 303099, *M. japonicum* MAFF303099; TONO, *M. loti* TONO; R7A, *M. japonicum* R7A; and 2037, *M. loti* NZP2037. Black and gray cells indicate the presence or absence of genes, respectively.

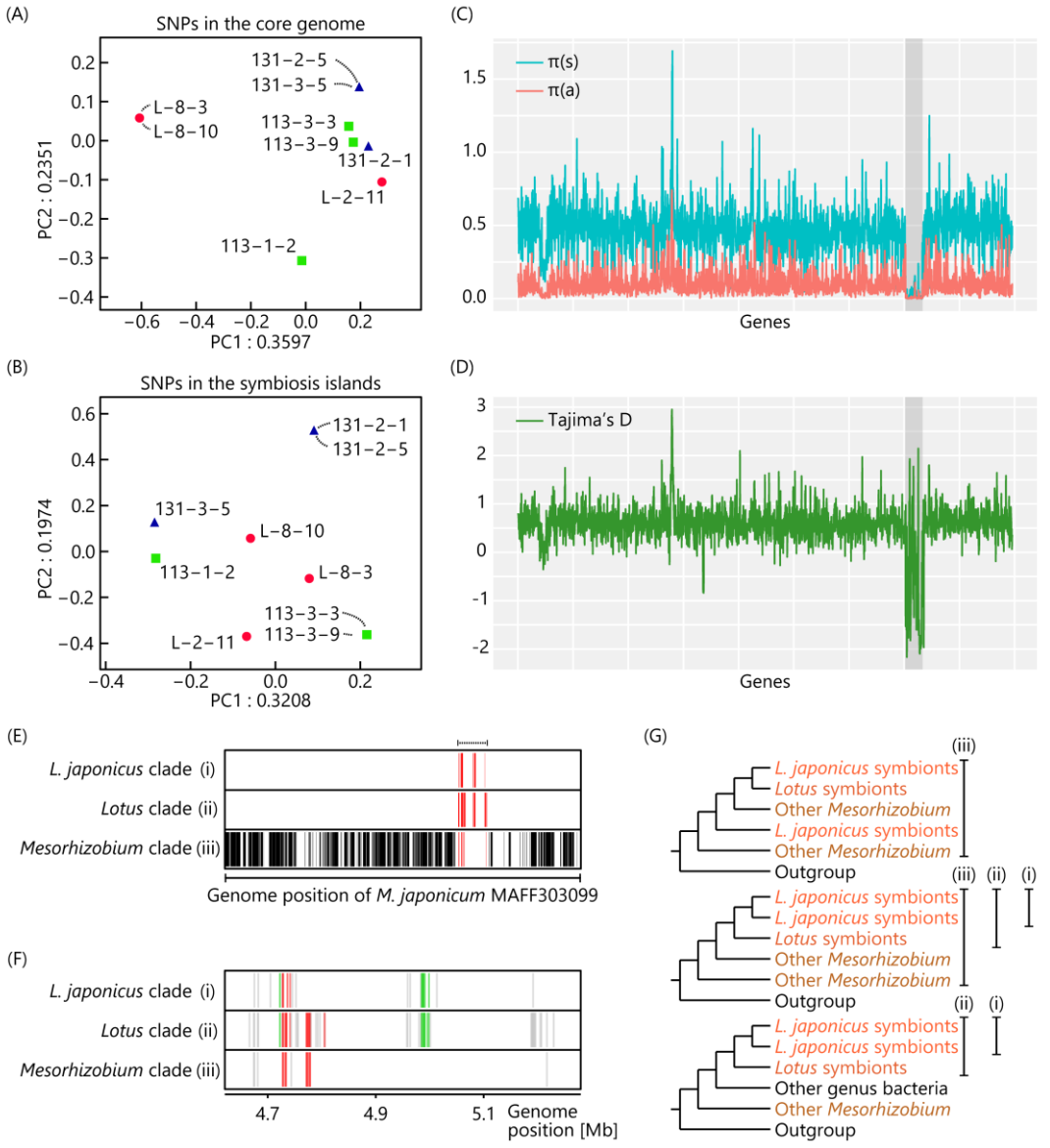

Figure 3. Genomic outlines of nine sequenced strains. (A, B) Principal component analyses of nine sequenced strains using single nucleotide polymorphisms (SNPs) in the core genome (A) and in the symbiosis island (B). Forms and colors of dots indicate sampling localities of each rhizobial symbiont with the strain type designated in brackets: blue triangle, Aomori (131); green square, Tottori (113); and red circle, Miyakojima (L). (C, D) Genome-wide distribution of nucleotide diversity statistics π (C) and Tajima's *D* (D), calculated for each single-copy gene described in Supporting

976 Information Table S4. The order of genes was based on the genome assembly of *M.*
977 *japonicum* MAFF303099. The shaded region indicates the symbiosis island. (C) Red
978 and blue lines indicate the nucleotide diversity statistic π at nonsynonymous and
979 synonymous sites, respectively. (E, F and G) Comparison of phylogenetic tree
980 topologies in the rhizobial whole genomes (E) and the symbiosis island (F). Bars are
981 indicated if: (i) *L. japonicus*-associated symbionts (nine sequenced strains, *M.*
982 *japonicum* MAFF303099 and *M. loti* TONO) form a single clade; (ii) *Lotus*-associated
983 symbionts (*L. japonicus*-associated symbionts and *M. loti* NZP2037, and *M. japonicum*
984 R7A) form a single clade; and (iii) *Mesorhizobium* strains (nine sequenced strains in
985 this study and other *Mesorhizobium* strains) form a single clade. (E) The black and red
986 bars indicate genes on the core genome and the symbiosis island, respectively. (F) Each
987 red, green and gray bar indicates nitrogen-fixing, nodulation and unknown genes. (G)
988 Schematic trees showing clades satisfying criteria i–iii listed above.

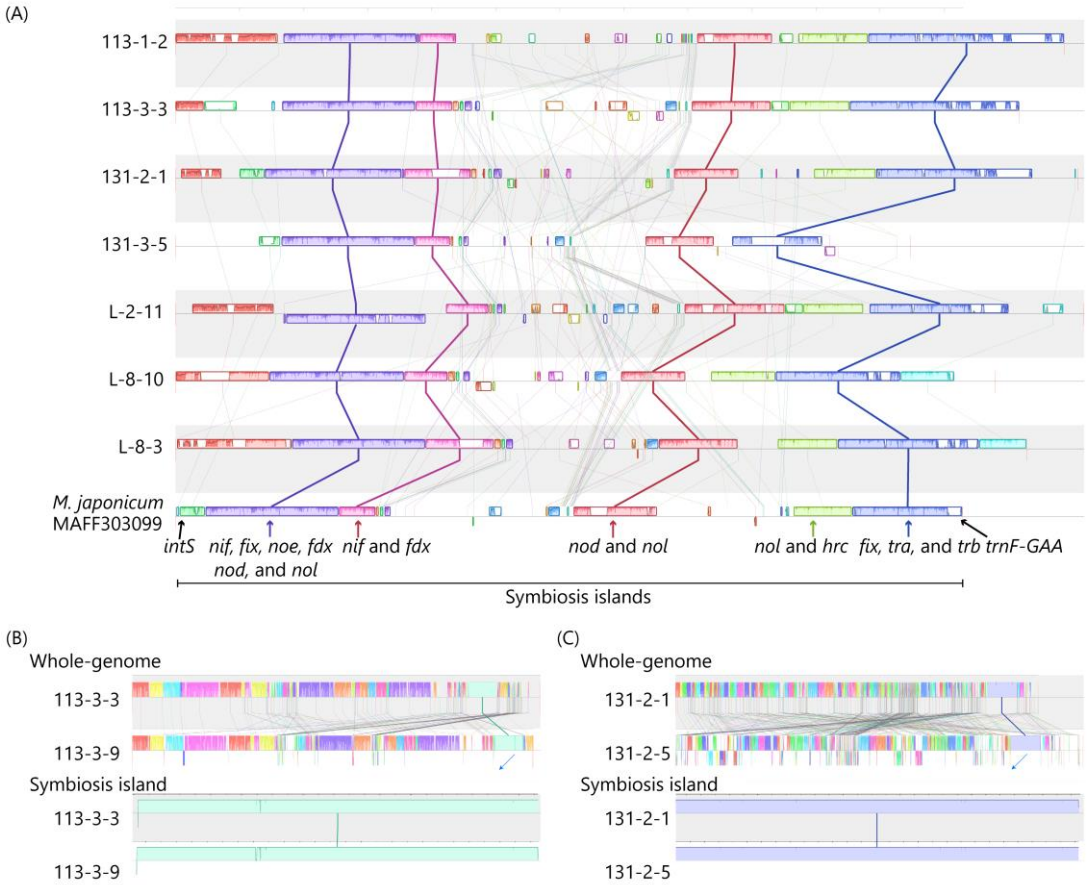

Figure 4. An alignment of the symbiosis islands of *L. japonicus*-associated symbionts, using progressiveMauve software. Bordered and connected boxes indicate similar sequence compositions among sequences. (A) The boxes connected by bold lines indicate conservative genetic clusters among all *L. japonicus*-associated symbionts. The purple and pink/purple conserved blocks, red conserved block and blue and yellow conserved blocks are referred as *nif–fdx* blocks, *nod* blocks and *nol–hrc* blocks in the main text, respectively. (B, C) Alignment of two pairs of rhizobial strains harboring almost identical symbiosis islands, 113-3-3 and 113-3-9 (B) and 131-2-1 and 131-2-5 (C), generated by progressiveMauve software. The upper and lower diagrams show alignments of whole genomes and the symbiosis islands, respectively. Boxes connected by bold lines indicate the symbiosis islands.

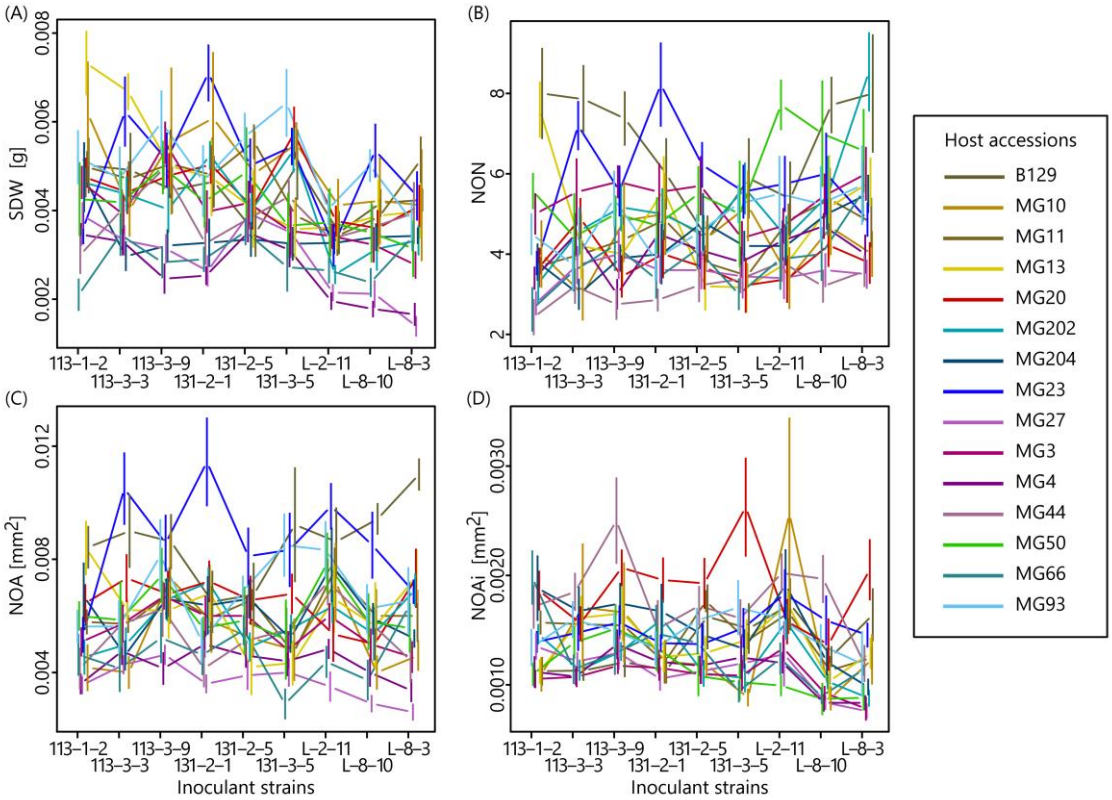

Figure 5. Plant phenotypes of the cross-inoculation experiments, (A) shoot dry weight (SDW), (B) nodule numbers (NON), (C) total nodule size (NOA), (D) nodule size per nodule (NOAi). Mean values and standard errors (bars) are shown.

SUPPLEMENTARY FIGURES

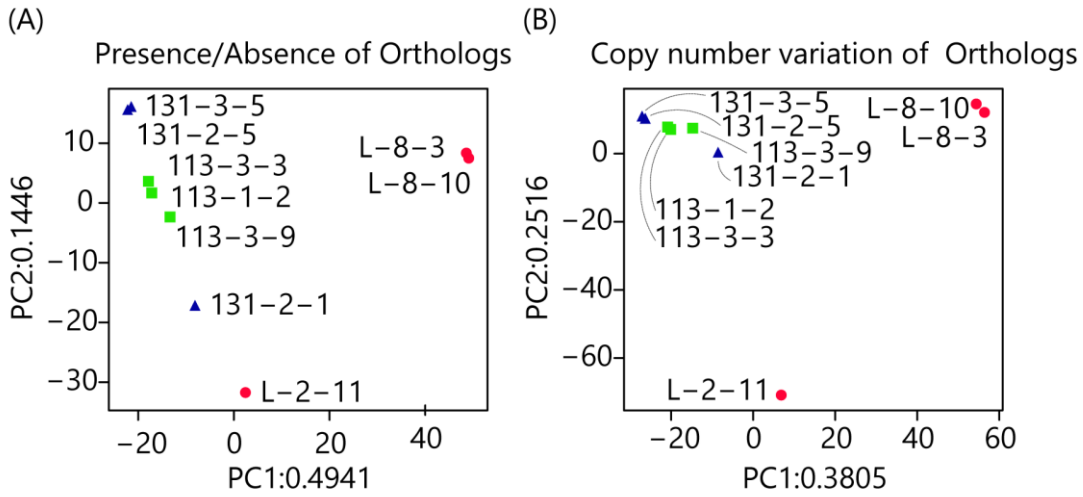

Figure S1. Principal component analysis (PCA) results of nine sequenced strains using ortholog profiles. (A) PCA based on the presence/absence of orthologs. (B) PCA based on copy number variations. Forms and colors of dots indicate the sampling localities of each rhizobial symbiont with the strain type designated in brackets: blue triangle, Aomori (131); green square, Tottori (113); and red circle, Miyakojima (L).

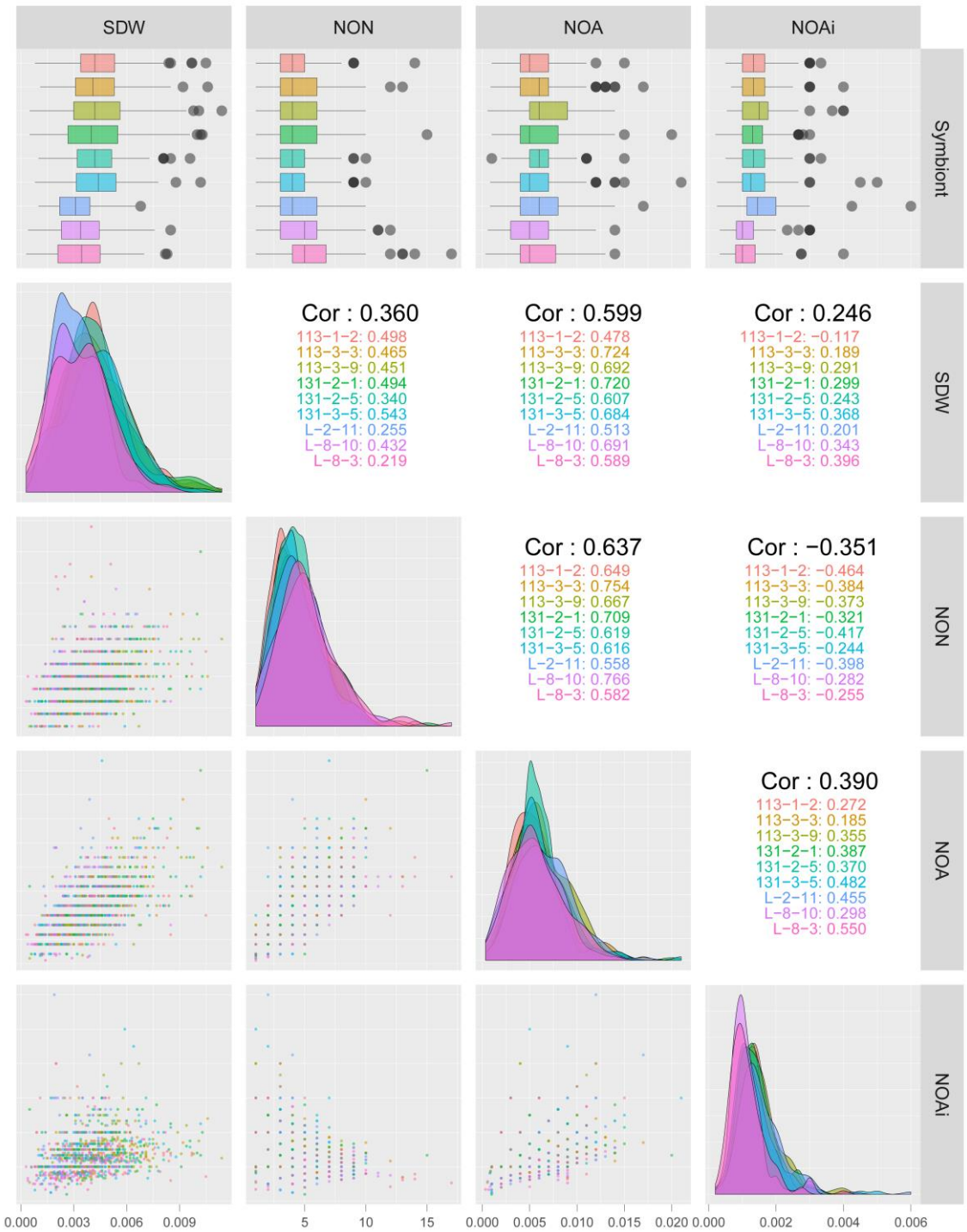

Figure S2. Correlations of phenotypes in rhizobial symbionts. Box plots show the distribution of phenotypic values for each rhizobial symbiont, and horizontal grey bars indicate the mean values. The x- and y-axes in each scatterplot represent phenotypic trait values: SDW, NON, NOA and NOAi. The diagonal plots are density curves for the

1053 individual points and Pearson's correlation coefficients are given for all phenotypes

1054 (Cor; black). Each color indicates one rhizobial symbiont.

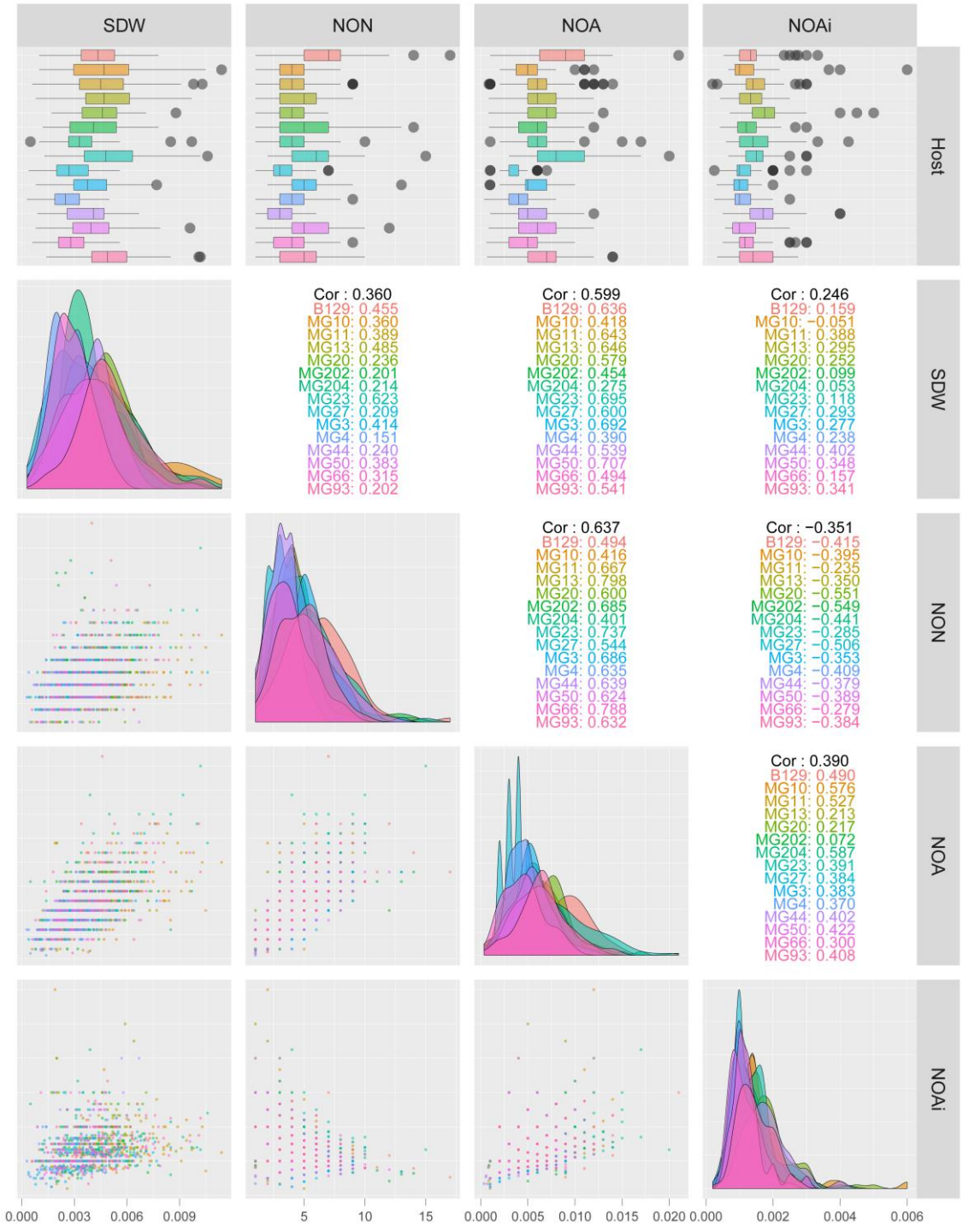

Figure S3. Correlations of phenotypes in *L. japonicus* accessions. Box plots show the distribution of phenotypic values for each accession, and horizontal grey bars indicate the mean values. The x- and y-axes in each scatterplot represent phenotypic trait values:

1074 SDW, NON, NOA and NOAi. The diagonal plots are density curves for the individual
1075 points and Pearson's correlation coefficients are given for all phenotypes (Cor; black).
1076 Each color indicates one *L. japonicus* accession.

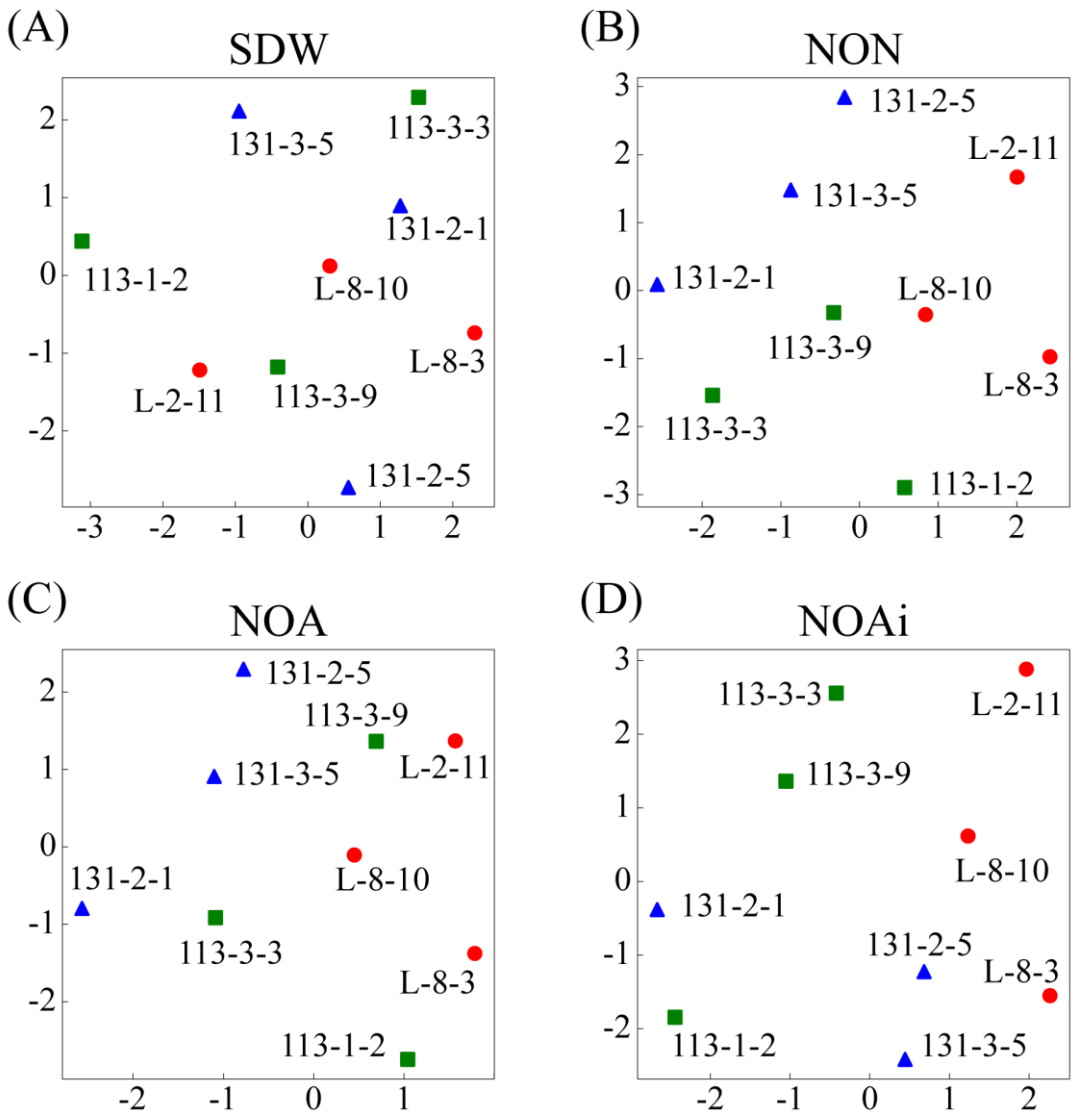

Figure S4. Non-metric multidimensional scaling of variations in G × G interactions of rhizobial symbionts: (A), SDW; (B), NON; (C), NOA; and (D), NOAi. Forms and colors of dots indicate sampling localities of each rhizobial symbiont with the strain type designated in brackets: blue triangle, Aomori (131); green square, Tottori (113); and red circle: Miyakojima (L).

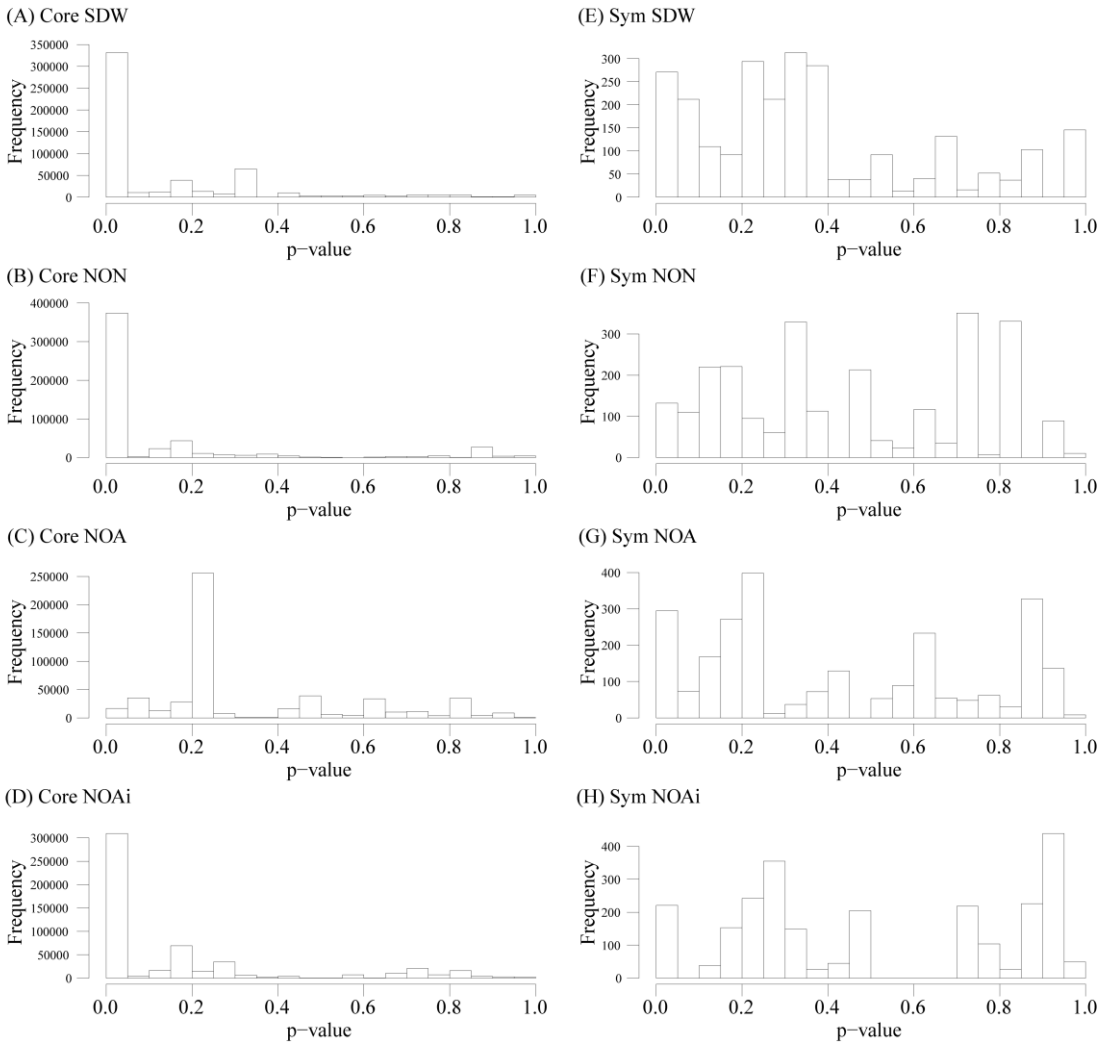

Figure S5. *P*-value distributions of correlations between variations in partner quality of each phenotype and the SNPs of the core genomes (A–D) and their symbiosis islands (E–H).

SUPPLEMENTAL FILES

Supplemental Table 1: Summary of de novo assembly

Supplemental Table 2: List of host accessions

Supplemental Table 3: List of reference genomes

Supplemental Table 4: Summary of next-generation sequencing data

Supplemental Table 5: List of single copy orthologs

Supplemental Table 6: Results of TukeyHSD test

Supplemental Table 7: Results of ANOVAs 1

Supplemental Table 8: Results of ANOVAs 2

Supplemental Table 9: Results of Mantel test 1

Supplemental Table 10: Results of Mantel test 2