AEM
Journals.ASM.org

Replicon-Dependent Differentiation of Symbiosis-Related Genes in *Sinorhizobium* Strains Nodulating *Glycine max*

Hui Juan Guo,[a,b,c] En Tao Wang,[d] Xing Xing Zhang,[a,b,c] Qin Qin Li,[a,b,c] Yan Ming Zhang,[a,b,c] Chang Fu Tian,[a,b,c] Wen Xin Chen[a,b,c]

State Key Laboratory of Agrobiotechnology, College of Biological Sciences, China Agricultural University, Beijing, China[a]; Key Laboratory of Soil Microbiology, Ministry of Agriculture, China Agricultural University, Beijing, China[b]; Rhizobium Research Center, China Agricultural University, Beijing, China[c]; Departamento de Microbiología, Escuela Nacional de Ciencias Biológicas, Instituto Politécnico Nacional, Mexico D.F., Mexico[d]

In order to investigate the genetic differentiation of *Sinorhizobium* strains nodulating *Glycine max* and related microevolutionary mechanisms, three housekeeping genes (*SMc00019*, *truA*, and *thrA*) and 16 symbiosis-related genes on the chromosome (7 genes), pSymA (6 genes), and pSymB (3 genes) were analyzed. Five distinct species were identified among the test strains by calculating the average nucleotide identity (ANI) of *SMc00019-truA-thrA*: *Sinorhizobium fredii*, *Sinorhizobium sojae*, *Sinorhizobium* sp. I, *Sinorhizobium* sp. II, and *Sinorhizobium* sp. III. These species assignments were also supported by population genetics and phylogenetic analyses of housekeeping genes and symbiosis-related genes on the chromosome and pSymB. Different levels of genetic differentiation were observed among these species or different replicons. *S. sojae* was the most divergent from the other test species and was characterized by its low intraspecies diversity and limited geographic distribution. Intergenic recombination dominated the evolution of 19 genes from different replicons. Intraspecies recombination happened frequently in housekeeping genes and symbiosis-related genes on the chromosome and pSymB, whereas pSymA genes showed a clear pattern of lateral-transfer events between different species. Moreover, pSymA genes were characterized by a lower level of polymorphism and recombination than those on the chromosome and pSymB. Taken together, genes from different replicons of rhizobia might be involved in the establishment of symbiosis with legumes, but these symbiosis-related genes might have evolved differently according to their corresponding replicons.

The genetic mechanism of adaptation to changing environmental conditions is a key topic in evolutionary biology (1). As a mutualistic symbiosis, rhizobia and their host legumes have long been recognized as an experimental model to investigate the chronic processes of infection of eukaryotic hosts by microbes (2). As symbiotic bacteria, two categories of genomic components, core and accessory genomes, have been reported (3, 4), and the symbiosis genes (*nif*, *nod*, and many other related genes) are a part of the accessory genome. However, the canonical *nodABC* genes are not required in some *Bradyrhizobium* strains for symbiosis with legumes (5, 6). Based on previous research into genetics and comparative genomics, it was hypothesized that no genes are common and specific to all the rhizobia (3, 7). In other words, the symbiotic abilities of different rhizobia have emerged independently, even though lateral gene transfer contributed to their evolution (3).

Glycine max (soybean) is one of the most important legume crops around the world. *Bradyrhizobium* and *Sinorhizobium* are recurrently reported as microsymbionts of soybeans (8–14). In contrast to the widely distributed *Bradyrhizobium*, *Sinorhizobium* strains nodulating soybeans were mainly found in Asia (9, 10, 15–18) and were reported as the dominant microsymbionts of soybeans in saline-alkaline soils (9, 10, 16, 17). Moreover, *Sinorhizobium* strains have been demonstrated to be successful soybean inoculants under these soil conditions (19, 20). Consistent with these findings, recent comparative genomics of soybean rhizobia revealed distinct genomic features specific to either *Sinorhizobium* or *Bradyrhizobium* (21). Some genes known to be involved in either alkaline-saline adaptations or symbiotic interactions were found to be specific to *Sinorhizobium* compared to *Bradyrhizobium* (21). This genomic analysis also highlighted the importance of studies on the microevolution of rhizobia. Recently, population genetics studies of *Bradyrhizobium* strains nodulating soybeans in South Asia and North America have been reported (11, 12) and further demonstrated that multilocus sequence analysis (MLSA) is a valuable tool in microevolutionary studies of rhizobia (22–24). The aim of this study was to reveal the genetic differentiation of *Sinorhizobium* strains nodulating *G. max*. This is of particular importance considering the great variations in symbiotic competence among *Sinorhizobium* strains nodulating soybeans (25). In this study, population genetics analyses were carried out for three housekeeping genes (*SMc00019*, *truA*, and *thrA*), six well-known symbiosis genes involved in regulation and biosynthesis of Nod factors (*nodD1*, *nodD2*, *nodC*, and *nodS* on pSymA and *nodM* and *nodE* on the chromosome), and 10 symbiosis-related genes involved in the optimization of rhizobium-legume interactions (*gcvT*, *purL*, *cobO*, *bacA*, and *cysD* on the chromosome; *rhcJ* and *y4wE* on pSymA; and *exoA*, *exoY*, and *phnC* on pSymB) (26–42). Genetic differentiation of the tested strains is discussed, considering the replicon origin of these housekeeping or symbiosis-related genes.

Received 10 September 2013 **Accepted** 30 November 2013
Published ahead of print 6 December 2013
Address correspondence to Chang Fu Tian, changfutian@gmail.com.
Supplemental material for this article may be found at http://dx.doi.org/10.1128/AEM.03037-13.

doi:10.1128/AEM.03037-13

MATERIALS AND METHODS

Strains. A total of 109 rhizobial strains used in this study were obtained from the Culture Collection of China Agricultural University (CCBAU) and were previously isolated from root nodules of *G. max* collected from 22 sites in 8 provinces of China (see Table S1 and Fig. S1 in the supplemental material). Geographically, these sites can be divided into two ecoregions corresponding to Xinjiang and the Huang Huai Hai Plain (or north China). Both ecoregions have saline-alkaline soils. Previously, these strains have been divided into five genomic groups: *Sinorhizobium fredii*, covering 79 strains; *Sinorhizobium* sp. I, consisting of 4 strains; *Sinorhizobium* sp. II, comprising 2 strains; and *Sinorhizobium* sp. III and *Sinorhizobium sojae*, with 12 strains each (9, 10, 43). This population composition reflects the real composition of *Sinorhizobium* in the root nodules of soybeans in the field (9, 10, 43). All the strains were grown in YMA medium at 28°C (44).

Gene amplification and sequencing. DNA from each strain was extracted as described previously (45) and used as the template for PCR amplification of 19 genes with homologs on the chromosome, a symbiosis plasmid (pSymA), or a chromid (pSymB) of *Sinorhizobium* sp. strain NGR234 (27, 28, 46). Gene locations and complete names are shown in Fig. S2 in the supplemental material. Primers and PCR protocols described previously were used for amplification of housekeeping core genes *SMc00019*, *truA*, and *thrA* (26). The primers for 16 symbiosis-related genes (see Table S2 in the supplemental material) were designed by using homologous regions in the genomes of *S. fredii* and *S. sojae* (21) using the software Primer 5.0. These symbiosis-related genes are *bacA*, *purL*, *nodM*, *gcvT*, *cysD*, *cobO*, and *nodE* on the chromosome (29–33); *nodC*, *nodS*, *y4wE*, *nodD1*, *nodD2*, and *rhcJ* on symbiosis plasmid pSymA (30, 31, 34–38); and *exoA*, *exoY*, and *phnC* on chromid pSymB (39–42). Subsequently, the PCR products were purified and commercially sequenced using an ABI 3730xl sequencer. The nucleotide sequences of each gene obtained were aligned and manually corrected using the program CLUSTAL W integrated in MEGA5 (47, 48).

Phylogenetic analyses. Neighbor-joining trees and maximum-likelihood (ML) trees were constructed using MEGA5 (48) and phyML (49), respectively, based on the models selected by the hierarchical likelihood ratio tests (hLRTs) of MODELTEST 3.7 (50). In order to reveal potentially incompatible signals in the evolutionary history, split phylogenetic networks (1,000 bootstraps) were inferred using the SPLITSTREE4 program (51). Concatenated sequences were also analyzed using CLONALFRAME (52) to infer the effect of recombination during the phylogenetic history. Five independent runs (100,000 burn-in iterations plus 200,000 sampling iterations for each run) were performed, and the run was judged satisfactory based on the previously described method integrated in CLONALFRAME (52, 53).

Nucleotide polymorphism and population genetics analyses. Statistics for nucleotide polymorphisms (the number of haplotypes [*h*], haplotype diversity [Hd], and nucleotide diversity [Pi]), population differentiation (fixation index [Fst]), and gene flow (number of migrants [Nm]) were estimated with DNASP v5 (54, 55). To investigate the admixture level of the *Sinorhizobium* populations in this study, the admixture LOCPRIOR model of STRUCTURE was used in analyses (56, 57). In this model, individual *i* inherited its genome from *K* "ancestral" subpopulations, and we used the discrete location information to assist in the clustering. If the "ancestry" of individuals is uncorrelated with the sampling locations, the model ignores the sampling information. Two criteria described previously were considered in choosing the ancestral population *K* (22). To line up the cluster labels across three independent STRUCTURE runs with the chosen *K*, the program CLUMPP was used (58). To estimate recombination within the populations, three analyses were performed: (i) the Shimodaira-Hasegawa (SH) test (59) for evaluation of the phylogenetic consistency among ML phylogenetic trees with the PAUP 4b10 program (60); (ii) calculation of minimal recombination events (Rm) with the DNASP program (54, 61); and (iii) calculation of breakpoints with the Recombination Detection Program (RDP) based on the concatenated data (62). CLONALFRAME (52) was used to calculate two recombination rate statistics: r/m (the relative impact of recombination compared with that of point mutation in the genetic diversification of the lineage) (63) and ρ/θ (the relative frequency of the occurrence of recombination compared with that of point mutation in the history of the lineage) (64).

Nucleotide sequence accession numbers. The 1,189 nucleotide sequences obtained in this study were deposited in the GenBank database under accession numbers KF381560 to KF381919, KF381980 to KF382699, KF816087 to KF816146, KF816032 to KF816034, KF816040 to KF816042, KF816046 to KF816050, KF816067 to KF816071, KF816073 to KF816076, KF816080 to KF816084, KF815983, KF815986, KF815990, KF815992, KF815997, KF816011, KF816014, KF816015, KF816024, KF816026, KF816027, KF816029, KF816038, KF816044, KF816052, KF816053, KF816056, KF816057, KF816059, KF816061, KF816063, KF816065, KF816078, and KF816086 (see Table S2 in the supplemental material).

RESULTS

Phylogenetic relationships of strains based upon genes from different replicons. In this study, the 109 strains were first grouped based on the phylogeny of *thrA* sequences (data not shown). The grouping results for the 109 strains were consistent with their previous species assignments. Then, 38 out of 79 strains of *S. fredii* were selected for further analysis by considering their *thrA* phylogenetic groups and sequence identity values (however, at least one strain was kept for each sampling site). Due to the relatively small number of strains available for *S. sojae* and *Sinorhizobium* sp. I, *Sinorhizobium* sp. II, and *Sinorhizobium* sp. III, all 30 strains were used for PCR amplification of the other 18 genes. Finally, the PCR products of these 18 genes were successfully obtained for 22/33 strains: *S. sojae*, 9 strains; *Sinorhizobium* sp. I, 3 strains; *Sinorhizobium* sp. II, 1 strain; and *Sinorhizobium* sp. III, 9 strains. Therefore, a subset of 60 strains (see Table S3 in the supplemental material) was used for further studies.

For each gene, the phylogenetic trees constructed by the neighbor-joining and ML methods showed similar topologies. In general, the grouping results were similar in the phylogenetic trees based on the concatenated sequences of the three core genes (*SMc00019*, *truA*, and *thrA*) (Fig. 1), the concatenated sequences of the seven symbiosis-related chromosomal genes (*bacA*, *purL*, *nodM*, *gcvT*, *cysD*, *cobO*, and *nodE*), and the concatenated sequences of three genes (*exoA*, *exoY*, and *phnC*) on pSymB (data not shown). In all three trees, the strains were well grouped according to their species, and some typing variations could be observed within *S. fredii* and *Sinorhizobium* sp. III (see Table S3 in the supplemental material). The phylogenetic relationships of the concatenated sequences for the six genes (*nodC*, *nodS*, *y4wE*, *nodD1*, *nodD2*, and *rhcJ*) on pSymA (Fig. 2) were different from those in the three trees mentioned above. In Fig. 2, the strains from different species were intermingled; diverse genotypes were found in *S. fredii* (10 types) and *Sinorhizobium* sp. III (5 types), the strains of *S. sojae* showed the same type as the dominant type of *S. fredii* strains, the only strain of *Sinorhizobium* sp. II showed a type identical with a minor type of *S. fredii*, and three strains of *Sinorhizobium* sp. I showed a unique type differing from the others.

Nucleotide diversity inferred from different genes. The intraspecies (96%) and interspecies (94%) average nucleotide identity (ANI) boundaries calculated from the three core genes (*SMc00019-truA-thrA*) clearly clustered the 60 representative strains into 5 species corresponding to *S. fredii*, *Sinorhizobium* sp. I, *Sinorhizobium* sp. II, *Sinorhizobium* sp. III, and *S. sojae*, as shown in Table S3 in the supplemental material. In nucleotide diversity

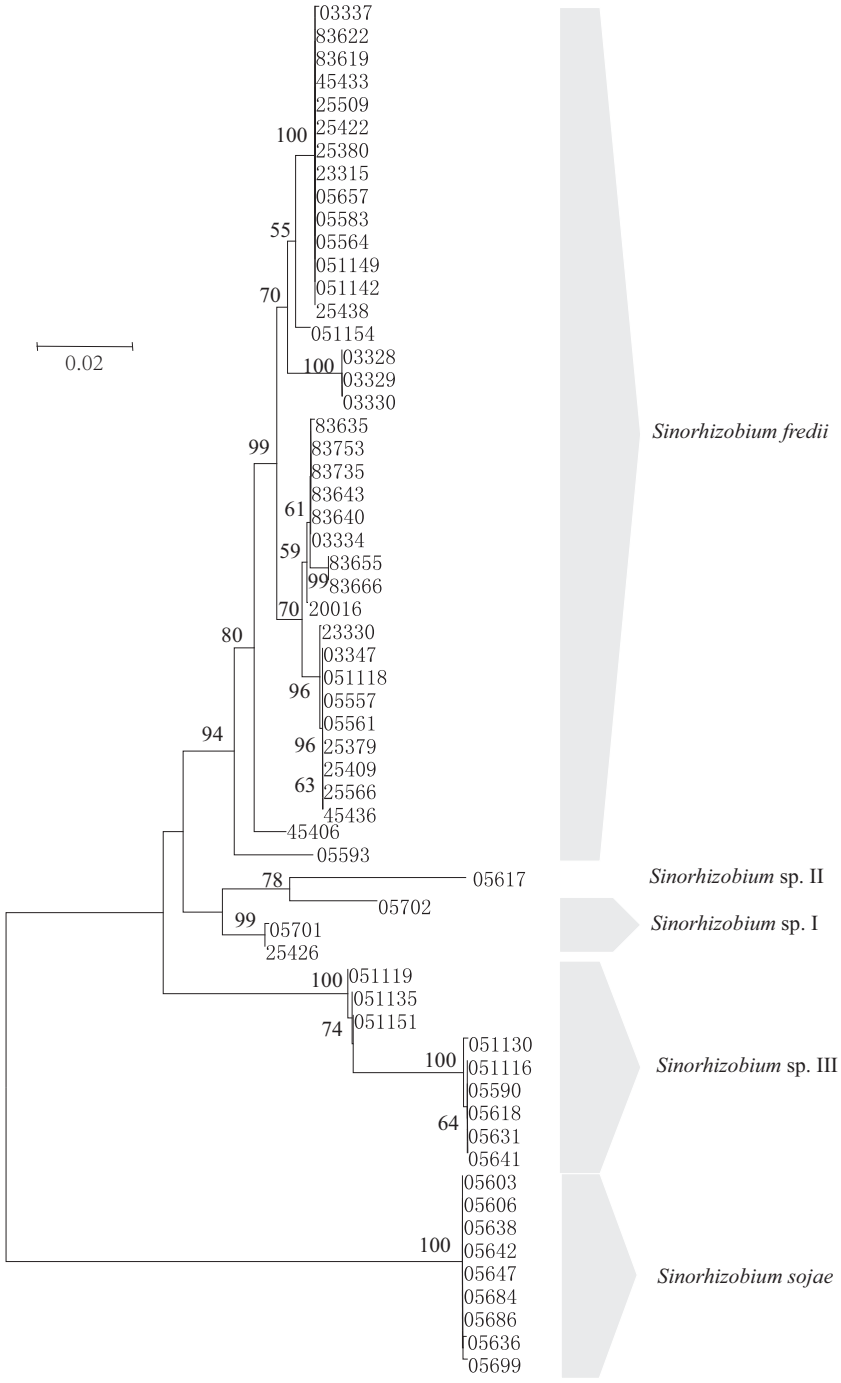

FIG 1 Neighbor-joining tree constructed based upon the concatenated sequences of core genes *SMc00019-truA-thrA*. The numbers are CCBAU strain numbers. Bootstrap values greater than 50% are indicated at the branch points. The scale bar represents 2% nucleotide substitutions.

analysis (Table 1), the nucleotide polymorphism statistics of core genes and symbiosis-related genes on the chromosome and pSymB were similar (averages for segregating sites, 93.4, 98.5, and 98.3, respectively; averages for *h* values, 11.3, 9.1, and 10.3; averages for Hd values, 0.788, 0.735, and 0.828; averages for Pi values, 0.06553, 0.05687 and 0.08348) and greater than those of the pSymA genes (segregating sites, 4.1; $h = 4.0$; Hd = 0.344; Pi = 0.00125).

In comparisons of the Hd and Pi values for different species (see Table S4 in the supplemental material), *S. fredii* and *Sinorhizobium* sp. III always showed apparently greater diversity (Hd = 0.583 to 0.893; Pi = 0.00112 to 0.02164) than *S. sojae* (Hd = 0.000 to 0.417; Pi = 0.00000 to 0.00053) in all the genes tested in this study. The values for *Sinorhizobium* sp. I and *Sinorhizobium* sp. II were not compared, since only a few strains were available in these groups.

Gene flow and genetic differentiation. In the estimations of housekeeping genes and symbiosis-related genes on the chromo-

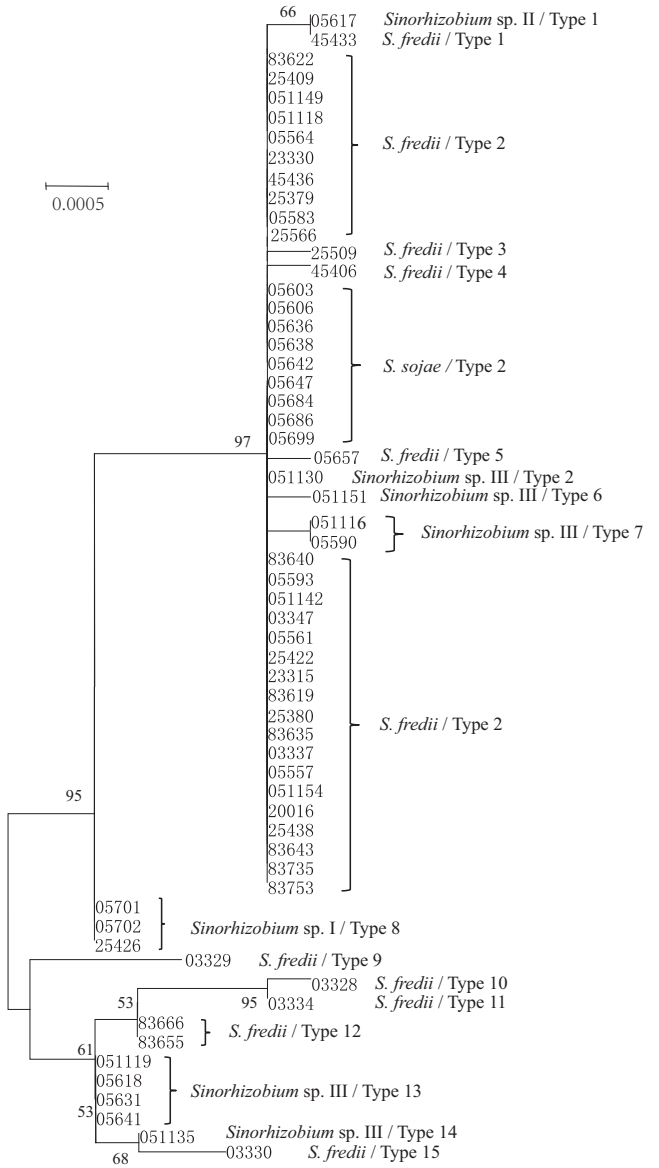

FIG 2 Neighbor-joining tree constructed based upon the concatenated sequences of the genes *nodC*, *nodS*, *y4wE*, *nodD1*, *nodD2*, and *rhcJ* on pSymA. The numbers are CCBAU strain numbers. Bootstrap values greater than 50% are indicated at the branch points. The scale bar represents 2% nucleotide substitutions.

some, pSymB, and pSymA (Table 2), the highest population differentiation (Fst) was found between *Sinorhizobium* sp. I and *S. sojae* (Fst = 0.76201 to 1.0) and the lowest Fst values (0.09399 to 0.14813) were detected between *S. fredii* and *Sinorhizobium* sp. I. However, the *P* values for all the Fst comparison pairs were statistically significant ($P < 0.05$), with the exception of the *S. fredii*-*S. sojae* pair of symbiosis-related genes on pSymA (Table 2). The Nm values, representing the frequency of genetic exchange, showed a tendency in contrast to that of the Fst values, as was expected (Table 2). The NeighborNet network of concatenated *SMc00019-thrA-truA* sequences of 60 representative strains (Fig. 3a) revealed a noticeable level of intraspecies genetic exchanges in *S. fredii*, *Sinorhizobium* sp. I, and *Sinorhizobium* sp. III. Gene flows were also observed among the lineages of *S. fredii*, *Sinorhizobium* sp. I and *Sinorhizobium* sp. II CCBAU 05617. However, *S. sojae* and *Sinorhizobium* sp. III were isolated from each other and showed rare gene flow with *S. fredii*. The trees of the NeighborNet network for the symbiosis-related genes on the chromosome and on pSymB showed topologies similar to that for the core genes (Fig. 3b and c). No obvious network could be detected in the NeighborNet tree of pSymA genes (Fig. 3d).

In STRUCTURE analyses, a *K* value of 4 was chosen for concatenated core genes (50,000 burn-in and 100,000 sampling iterations), a *K* value of 4 for symbiosis-related chromosomal genes (100,000 burn-in and 1,000,000 sampling iterations), a *K* value of 4 for pSymB genes (10,000 burn-in and 100,000 sampling iterations), and a *K* value of 2 for pSymA genes (50,000 burn-in and 1,000,000 sampling iterations). The results for concatenated core genes (Fig. 4a) and for symbiosis-related genes on the chromosome (Fig. 4b) and on pSymB (Fig. 4c) were similar, while pSymA genes showed a quite different structure pattern (Fig. 4d). In Fig. 4a to c, *S. sojae* strains were well differentiated from the other species; *Sinorhizobium* sp. III was distinct from the other species, with the exception of a low level of gene flow with *S. fredii*; and *Sinorhizobium* sp. I showed a high admixture level with *Sinorhizobium* sp. II CCBAU 05617 and received noticeable genetic information from *S. fredii* or vice versa. As shown in Fig. 4d, pSymA

TABLE 1 Nucleotide polymorphism of genes

Gene	Size (bp)	No. of segregating sites	*h*	Hd	Pi
Core genes					
SMc00019	373	91	11	0.797	0.07265
thrA	597	127	12	0.844	0.06809
truA	308	63	11	0.731	0.05585
Concatenate	1,278	281	23	0.914	0.06647
Avg	426.0	93.4	11.3	0.788	0.06553
Symbiosis-related genes on chromosome					
bacA	383	184	6	0.562	0.10951
purL	425	89	10	0.833	0.00050
nodM	440	116	15	0.837	0.08151
gcvT	351	83	9	0.752	0.06117
cysD	460	71	7	0.623	0.04301
cobO	410	67	7	0.724	0.05032
nodE	456	80	10	0.815	0.05208
Concatenate	2,925	690	25	0.936	0.06484
Avg	417.8	98.5	9.1	0.735	0.05687
Symbiosis-related genes on pSymB					
exoA	322	103	8	0.793	0.09668
exoY	303	84	10	0.821	0.08247
phnC	463	108	13	0.869	0.07130
Concatenate	1,088	295	19	0.894	0.08213
Avg	362.7	98.3	10.3	0.828	0.08348
Symbiosis-related genes on pSymA					
nodC	572	2	2	0.364	0.00127
nodS	403	5	4	0.378	0.00149
y4wE	493	6	6	0.405	0.00194
nodD2	432	9	7	0.463	0.00200
nodD1	541	1	2	0.033	0.00006
rhcJ	438	2	3	0.332	0.00077
Concatenate	2,879	25	15	0.596	0.00122
Avg	479.8	4.1	4.0	0.344	0.00125

TABLE 2 Fst and Nm values between each pair of species

Species	Value[a]			
		Sinorhizobium sp.		
	S. fredii	I	III	*S. sojae*
Housekeeping genes				
S. fredii		0.2	0.08	0.02
Sinorhizobium sp. I	0.09399[d]		0.11	0.03
Sinorhizobium sp. III	0.25514[d]	0.34535[d]		0.02
S. sojae	0.36543[d]	0.76201[d]	0.66570[d]	
Symbiosis-related genes on chromosome				
S. fredii		0.03	0.06	0.02
Sinorhizobium sp. I	0.13499[d]		0.03	0.00
Sinorhizobium sp. III	0.26549[d]	0.41043[b]		0.01
S. sojae	0.36791[d]	0.79999[c]	0.65599[c]	
Symbiosis-related genes on pSymB				
S. fredii		0.07	0.07	0.03
Sinorhizobium sp. I	0.11360[d]		0.00	0.00
Sinorhizobium sp. III	0.27607[d]	0.74659[b]		0.00
S. sojae	0.34241[d]	1.00000[d]	0.88609[c]	
Symbiosis-related genes on pSymA				
S. fredii		0.28	1.98	4.61
Sinorhizobium sp. I	0.14813[d]		0.73	0.00
Sinorhizobium sp. III	0.13724[d]	0.13835[b]		0.59
S. sojae	−0.01323[e]	1.00000[d]	0.34374[c]	

[a] The numbers are Nm values in the upper triangles and Fst in the lower triangles.
[b] $0.01 < P < 0.05$.
[c] $0.001 < P < 0.01$.
[d] $0.0001 < P < 0.01$.
[e] Nonsignificant.

genes seemed to have two ancestors. The dominant one contributed to the pSymA gene pool of all five *Sinorhizobium* species, whereas the other was found in some strains of *S. fredii* (6/38), *Sinorhizobium* sp. III (5/9), and *Sinorhizobium* sp. I (3/3). Notably, pSymA genes of *Sinorhizobium* sp. I were identified as recombinants of two ancestors. However, the patterns observed for *Sinorhizobium* sp. I and particularly for *Sinorhizobium* sp. II need further investigation when more strains are available for the two species.

Recombination in the history of the *Sinorhizobium* lineages. By DNASP analysis, 17, 10, 13, and 42 putative recombination events (Rm) were found in *thrA*, *SMc00019*, *truA*, and their concatenated sequences, respectively. Notably (Table 3), Rm values were noticeably lower in symbiosis-related pSymA genes (Rm = 1) than in chromosome (Rm = 114) and pSymB (Rm = 38) genes, implying that recombination in the chromosome and pSymB was more frequent than in test pSymA genes. The greater role of recombination in the evolution of housekeeping genes and symbiosis-related genes on the chromosome and pSymB than in that of pSymA genes was also confirmed by CLONALFRAME analysis (Table 3). For concatenated sequences, r/m and ρ/θ were 10.7 and 0.69 for *SMc00019-thrA-truA*, 1.13 and 0.06 for symbiosis-related chromosomal genes, 10.9 and 0.96 for pSymB, and 0.62 and 0.34 for pSymA. In the SH test (Table 4), all gene trees were significantly incongruent with the inferred species phylogeny based on *SMc00019-thrA-truA* ($P < 0.05$).

DISCUSSION

***Sinorhizobium* strains nodulating soybeans belong to five species.** Previously, three housekeeping genes (*SMc00019-truA-thrA*) were demonstrated to be able to reflect the intraspecific and interspecific genomic differences among the rhizobial strains and were suggested as taxonomic markers for differentiating the rhizobial species (26). In the present study, ANI analyses of these three markers grouped the 60 representative strains nodulating soybeans into five species: *Sinorhizobium* sp. I, *Sinorhizobium* sp. II, *Sinorhizobium* sp. III, *S. sojae*, and *S. fredii* (Fig. 1; see Table S3 in the supplemental material). These results are consistent with earlier clustering results for *Sinorhizobium* strains nodulating soybeans (9, 10, 14, 43), though only *S. sojae* was formally proposed as a different species from *S. fredii* (14). In Fst analysis of these three housekeeping genes (Table 2), *S. sojae* was found to be the most divergent species. Similarly, *Sinorhizoibum* sp. III showed very great genetic differentiation from *S. fredii*, *Sinohirozibum* sp. I, and *S. sojae* (Fst above 0.25). Moreover, these Fst values were all statistically significant ($P < 0.001$), indicating robust species assignment by ANI of *SMc00019-truA-thrA* in this study. In line with the Fst values, different levels of differentiation between *Sinorhizobium* species were also revealed by NeighborNet tree and STRUCTURE analyses, as described above (Fig. 3a and 4a).

Interestingly, notably lower nucleotide polymorphism (Hd and Pi values) was found in *S. sojae* than in *Sinorhizobium* sp. III and *S. fredii* (at least nine strains for each species were available in this study). The underlying mechanisms remain elusive. However, the biogeographic patterns of these species might provide some interesting hints. In contrast to the wider distribution of *S. fredii* and *Sinorhizobium* sp. III, *S. sojae* was found only in a small part of Hebei Province in China (see Table S3 in the supplemental material) (9, 10, 43). Moreover, the genome size of *S. sojae* strain CCBAU 05684 (5.97 Mb) was smaller than those of *S. fredii* (6.74 ± 0.14 Mb) and *Sinorhizobium* sp. III strain CCBAU 05631 (6.30 Mb) (21). Therefore, it is probable that *S. sojae* was well adapted to certain local regions of Hebei Province while *S. fredii* and *Sinorhizobium* sp. III had the ability to adapt to more diverse environmental conditions. Could geographic isolation play a role in the differentiation between *S. sojae* and *S. fredii* or *Sinorhizobium* sp. III? Fst analyses suggested no geographic isolation between *S. fredii* populations in Xinjiang and the Huang Huai Hai region (data not shown; see Fig. S1 and Table S3 in the supplemental material), despite more than 2,000 km between the two regions. Furthermore, *S. fredii* and *Sinorhizobium* sp. III were also found in the sampling sites where *S. sojae* strains were isolated (see Table S3 in the supplemental material). Interestingly, the lower diversity of *Sinorhizobium medicae* than of *Sinorhizobium meliloti* (both species nodulate *Medicago* spp.) was recurrently observed in the same sampling sites (65, 66). Even though they can cooccur in sympatry, significant differentiation of chromosomal loci between *S. medicae* and *S. meliloti* was observed (67). This suggested a rather ancient speciation event leading to the two species (67). A similar hypothesis could apply to the speciation between *S. sojae* and *S. fredii*.

Replicon-dependent evolution of symbiosis-related genes. It is well known that *nod* and *nif* genes are the key players in most rhizobium-legume symbiotic systems. These genes, especially *nod*

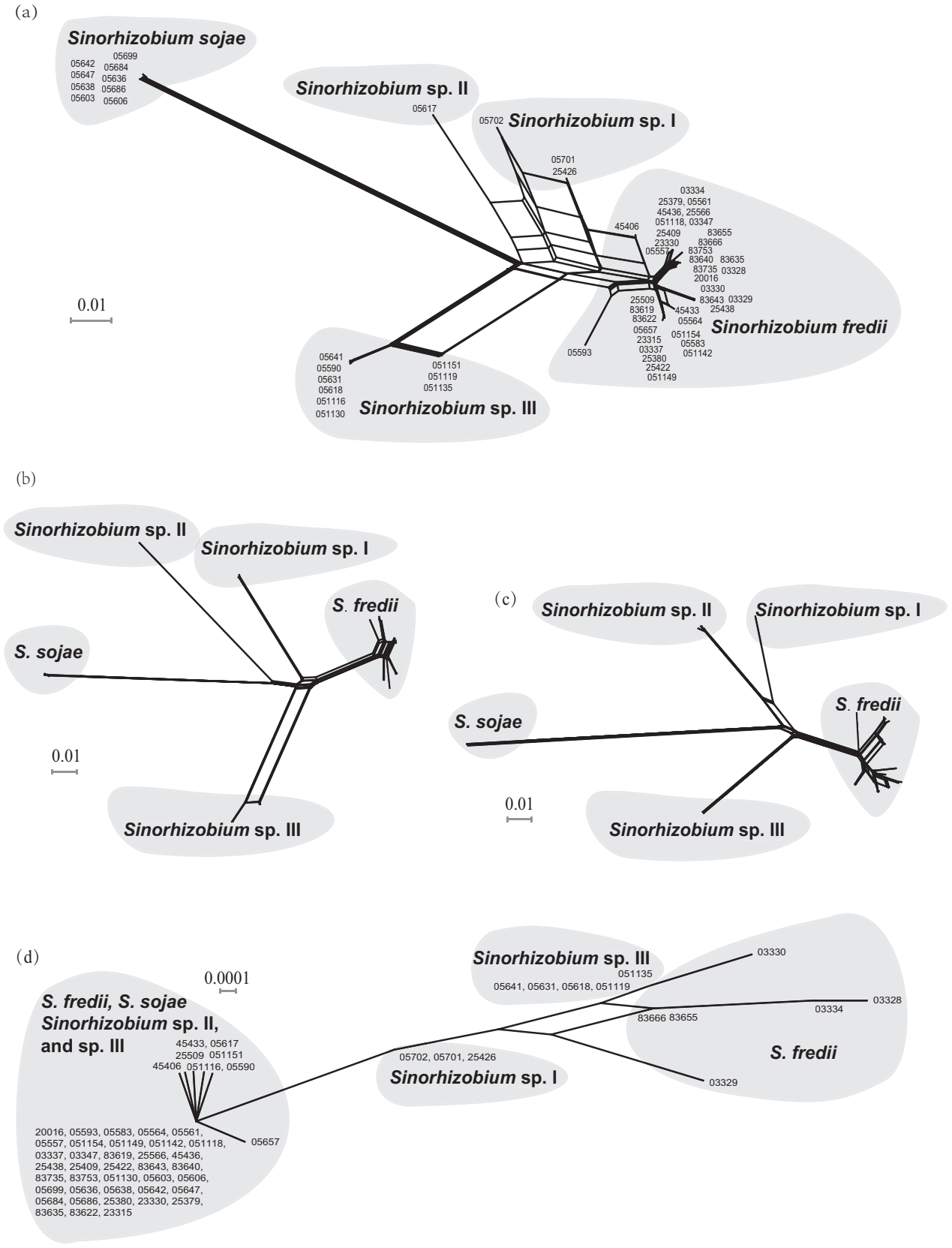

FIG 3 NeighborNet network trees. (a) Concatenated *SMc00019-thrA-truA* sequences. (b) Symbiosis-related genes on the chromosome (*bacA-purL-nodM-gcvT-cysD-cobO-nodE*). (c) Symbiosis-related genes on pSymB (*exoA-exoY-phnC*). (d) Symbiosis-related genes on pSymA (*nodC-nodS-y4wE-nodD1-nodD2-rhcJ*). The numbers are the CCBAU strain numbers. Strain numbers are not shown in panels b and c due to space limitations but are available upon request.

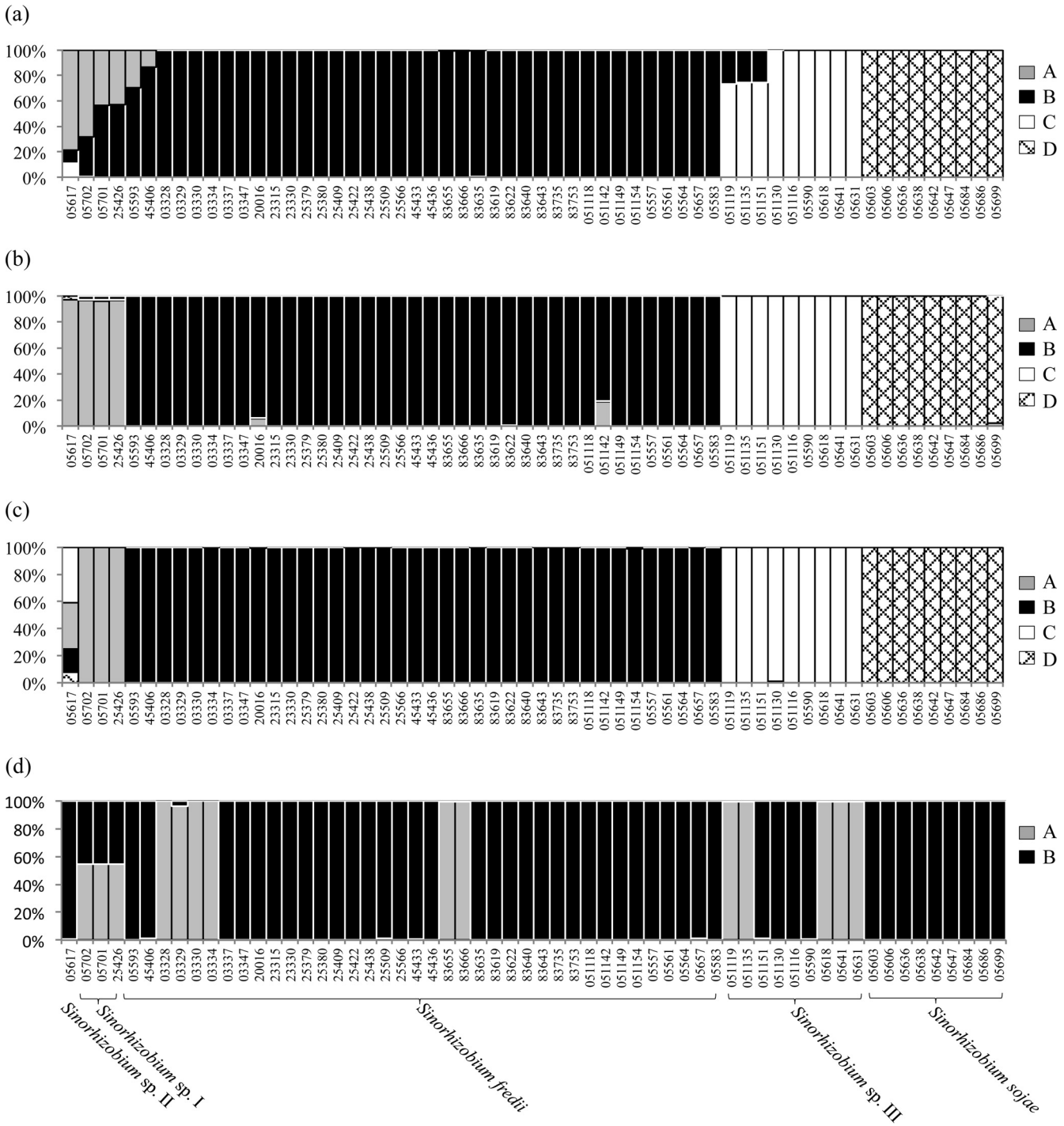

FIG 4 STRUCTURE analyses of *Sinorhizobium* populations. Housekeeping genes *SMc00019-thrA-truA* (a) and symbiosis-related genes on the chromosome (b), pSymB (c), and pSymA (d) were analyzed. The inferred ancestries are named A to D (a to c) or A and B (d). The horizontal axis represents current *Sinorhizobium* individuals (in the same order in all panels), and the bar for each individual is filled according to the inferred proportions of single-nucleotide alleles that were derived from each of the ancestries.

genes, which are overrepresented in rhizobia, are called typical symbiosis genes (7). Many other genes performing certain general functions were also essential for symbiosis to function well (3, 7, 31). These symbiosis-related genes are now the major topics of many studies on rhizobium-legume interactions (3, 21, 68). However, the evolution of these symbiosis-related genes, other than the *nod* and *nif* genes, is largely unknown. *Sinorhizobium* genomes are characterized by the presence of three main replicons (a chromosome; a chromid, pSymB; and a megaplasmid, pSymA), and *nif* and most *nod* genes are located on pSymA (68–71). In this study,

TABLE 3 Recombination analysis by DNASP and CLONALFRAME

Gene	Size (bp)	Rm[a]	r/m[b]	ρ/θ[c]
Concatenated symbiosis-related genes				
Chromosome	2,925	114	1.13	0.06
pSymA	2,879	1	0.62	0.34
pSymB	1,088	38	10.9	0.96
Concatenated core genes				
SMc00019-thrA-truA	1,278	42	10.7	0.69

[a] Rm, observed minimum number of recombination events (61).
[b] r/m, the relative impact of recombination compared with that of point mutation in the genetic diversification of the lineage.
[c] ρ/θ, the relative frequency of the occurrence of recombination compared with that of point mutation in the history of the lineage.

TABLE 4 SH analysis of each gene locus in comparison with the concatenated core genes

Tree	$-\ln L$[a]	Diff $-\ln L$[b]	P[c]
thrA	4,612.17	594.29	0.000
SMc00019	5,022.77	1,004.89	0.000
truA	5,571.46	1,553.58	0.000
bacA	5,343.32	1,381.79	0.000
cobO	4,824.70	863.17	0.002
cysD	5,184.10	1,222.56	0.000
gcvT	9,045.21	5,083.68	0.000
nodM	4,996.29	1,034.76	0.000
nodE	4,707.38	745.84	0.004
purl	4,621.62	660.08	0.020
nodC	9,796.29	5,834.75	0.000
nodD1	9,806.13	5,844.60	0.000
nodD2	9,694.03	5,732.49	0.000
y4wE	9,416.71	5,455.18	0.000
nodS	9,796.84	5,835.30	0.000
rhcJ	9,787.48	5,825.95	0.000
exoA	4,766.20	804.67	0.003
exoY	4,847.46	885.92	0.002
phnC	4,641.59	618.02	0.000

[a] $-\ln L$, negative log-likelihood value for the constrained topology.
[b] Diff $-\ln L$, score difference between the nonconstrained and constrained trees.
[c] Significance of the difference in $-\ln L$ scores achieved by the constrained and unconstrained trees as assessed by the SH test.

we investigated the evolution of *nod* genes involved in regulation (*nodD1* and *nodD2* on pSymA), biosynthesis of the glucosamine (chitin) oligosaccharide backbone (*nodC* on pSymA and *nodM* on the chromosome), and modification (*nodS* on pSymA) and biosynthesis of fatty acyl at the nonreducing termini (*nodE* on the chromosome) of Nod factors (72). In contrast to other tested *nod* genes on pSymA, it was found that *nodE* and *nodM* showed evolutionary characteristics similar to those of other housekeeping genes or symbiosis-related genes on the chromosome. The requirement for chromosomal genes, such as *gcvT*, in the modulation of cultivar-specific symbiosis with soybeans further implies that the optimization of symbiosis efficiency could be a long-term evolutionary process (3, 32, 73, 74). *y4wE* on pSymA, encoding a tryptophan transferase involved in indole-3-acetic acid (IAA) synthesis, is essential for the flavonoid-inducible IAA production of *Sinorhizobium* sp. NGR234 (38). An increased level of IAA has been found to be good for rhizobium-legume symbiosis (37, 75, 76). In this study, the evolution pattern of *y4wE* was found to be similar to that of *rhcJ* and *nod* genes on pSymA. In addition to these examples, the other tested symbiosis-related genes could also be grouped into chromosome-pSymB and pSymA categories by using various population genetics estimates used in this study (Table 1). These findings suggested replicon-dependent evolution of tested symbiosis-related genes.

Moreover, the levels of genetic differentiation among *Sinorhizobium* species nodulating soybeans were found to vary for symbiosis-related genes on different replicons (Table 2). Symbiosis-related genes on the chromosome and pSymB showed very great genetic differentiation (Fst $>$ 0.25; $P < 0.01$) among *S. sojae*, *S. fredii*, and *Sinorhizobium* sp. III. Symbiosis-related genes of pSymA were very greatly differentiated in *S. sojae-Sinorhizobium* sp. I and *S. sojae-Sinorhizobium* sp. III pairs (Fst $>$ 0.25; $P < 0.01$) but moderately differentiated in the pairs *S. fredii-Sinorhizobium* sp. I, *S. fredii-Sinorhizobium* sp. III, and *Sinorhizobium* sp. I-*Sinorhizobium* sp. III ($0.05 <$ Fst < 0.15; $P < 0.05$). No significant differentiation of pSymA genes was observed between *S. fredii* and *S. sojae*. The high Nm values observed in pSymA data (Table 2) and the results of STRUCTURE analyses (Fig. 4) suggested the lateral transfer of pSymA genes or fragments in *Sinorhizobium* populations. Therefore, these population genetics findings may explain the paraphyletic group of *S. fredii* strains intermingled with other species in Fig. 2 and 3d. However, strong evidence for the lateral transfer of pSymA genes was observed only for the dominant type 2, which is harbored by *S. fredii*, *S. sojae*, and *Sinorhizobium* sp. III (Fig. 2). Further investigations of the host specificity of this dominant type 2 and other pSymA types may uncover the biological functions of these divergences in pSymA sequences, as reported for *S. meliloti* and *S. medicae* strains nodulating *Medicago* spp. (67).

The contribution of recombination to genetic diversity has been reported to vary in different rhizobial populations, such as *Rhizobium leguminosarum*, *Sinorhizobium*, and *Bradyrhizobium* populations (11, 12, 21, 22, 24, 66). Here, extensive intergenic recombination was revealed by SH testing of single-gene phylogeny and species phylogeny based on core genes. The incongruences between species phylogeny and the phylogenies of symbiosis-related genes on the chromosome or pSymB were largely caused by intraspecies recombination rather than interspecies recombination (Fig. 3a to c and 4). In contrast, symbiosis-related genes on pSymA might be frequently transferred between different species (Fig. 3d and 4). Interestingly, no recombination breakpoints were found in pSymA genes by using RDP (62), suggesting their similar evolutionary histories (see Table S5 in the supplemental material). An Rm value of 1 and an r/m value of 0.62 further indicated a limited role of recombination in test symbiosis-related genes on pSymA compared to those on pSymB (Rm = 38 and r/m = 10.9) and the chromosome (Rm = 114 and r/m = 1.13). However, pSymA genes might not be immune to recombination, as pSymA genes of *Sinorhizobium* sp. I were found to be recombinants of two ancestors inferred by STRUCTURE analysis (Fig. 4d). Is there any bias of recombination in different replicons? In analyses of nucleotide polymorphism (Table 1), similar values of segregating sites, number of haplotypes, haplotype diversity, and nucleotide diversity among the core genes and symbiosis-related genes on the chromosome and pSymB were observed. These values were at least two times greater than those for the pSymA genes (Table 1). Lower diversity and fewer recombination

events in the symbiotic plasmid than in the chromosomal genes were also reported in other rhizobia, such as *Rhizobium etli* (77).

In line with the lower polymorphism of symbiosis-related genes on pSymA than in those on the chromosome and pSymB, only two ancestral populations were found for pSymA genes, while four ancestral populations were revealed for symbiosis-related genes on pSymB and the chromosome (Fig. 4). Moreover, the two ancestral populations of pSymA genes showed random geographic distribution in this study (Fig. 4d; see Table S3 in the supplemental material). These patterns implied that symbiosis-related genes on pSymA (*nodC*, *nodS*, *y4wE*, *nodD1*, *nodD2*, and *rhcJ*) might have been subjected to severe selection by the host (soybean) and were consistent with the ability of pSymA to shuttle freely between populations. In line with this view, these pSymA genes showed noticeably lower G+C percentages (56.49% ± 1.82% [average ± standard deviation]) than the pSymB (63.49% ± 2.32%) and chromosomal (63.22% ± 3.84%) genes and perfectly meet the criteria for accessory genes (4, 68). On the other hand, symbiosis-related genes on pSymB (*exoA*, *exoY*, and *phnC*) showed levels of nucleotide diversity, genetic differentiation, and recombination similar to those of genes on the chromosome (*bacA*, *purL*, *nodM*, *gcvT*, *cysD*, *cobO*, and *nodE*) and core genes (*SMc00019-truA-thrA*). This further supported the chromid theory of a pSymB-like plasmid, where the chromid genes ameliorate toward the nucleotide composition and codon usage of the chromosomal genes when the chromid is taking more essential functions in the long-term evolution (46, 71). It was hypothesized that the chromid could be a defining feature of a genus (46), and recent comparative genomics studies and observations in this study also suggested that the chromid has a distinct role in intragenus and intraspecies differentiation (69).

In conclusion, five species were defined among the *Sinorhizobium* strains nodulating soybeans. Different levels of genetic differentiation were observed among these species or different replicons. *S. sojae* was most divergent from the other test species and was characterized by its low intraspecies diversity and limited geographic distribution. There was no geographic isolation between *S. fredii* populations in different ecoregions in China. Although intergenic recombination dominated the evolution of 19 genes from different replicons, recombination happened frequently among strains within the same species but rarely between different species in core genes and symbiosis-related genes on the chromosome and pSymB. pSymA genes showed a clear pattern of lateral transfer events between different species and were characterized by a lower level of polymorphism and recombination than those on the chromosome and pSymB. Taken together, gene functions from different replicons of rhizobia might be integrated with each other to establish symbiosis with legumes, but these symbiosis-related genes might have evolved differently according to their corresponding replicons.

ACKNOWLEDGMENTS

This study was supported by the National Natural Science Foundation of China (31200002) and National Basic Research Program of China (973 Program) grant 2010CB126500.

REFERENCES

1. **Stapley J, Reger J, Feulner PGD, Smadja C, Galindo J, Ekblom R, Bennison C, Ball AD, Beckerman AP, Slate J.** 2010. Adaptation genomics: the next generation. Trends Ecol. Evol. **25:**705–712. http://dx.doi.org/10.1016/j.tree.2010.09.002.
2. **Batut J, Andersson SG, O'Callaghan D.** 2004. The evolution of chronic infection strategies in the alpha-proteobacteria. Nat. Rev. Microbiol. **2:**933–945. http://dx.doi.org/10.1038/nrmicro1044.
3. **Masson-Boivin C, Giraud E, Perret X, Batut J.** 2009. Establishing nitrogen-fixing symbiosis with legumes: how many rhizobium recipes? Trends Microbiol. **17:**458–466. http://dx.doi.org/10.1016/j.tim.2009.07.004.
4. **Young JPW, Crossman LC, Johnston AW, Thomson NR, Ghazoui ZF, Hull KH, Wexler M, Curson AR, Todd JD, Poole PS, Mauchline TH, East AK, Quail MA, Churcher C, Arrowsmith C, Cherevach I, Chillingworth T, Clarke K, Cronin A, Davis P, Fraser A, Hance Z, Hauser H, Jagels K, Moule S, Mungall K, Norbertczak H, Rabbinowitsch E, Sanders M, Simmonds M, Whitehead S, Parkhill J.** 2006. The genome of *Rhizobium leguminosarum* has recognizable core and accessory components. Genome Biol. **7:**R34. http://dx.doi.org/10.1186/gb-2006-7-4-r34.
5. **Giraud E, Moulin L, Vallenet D, Barbe V, Cytryn E, Avarre JC, Jaubert M, Simon D, Cartieaux F, Prin Y, Bena G, Hannibal L, Fardoux J, Kojadinovic M, Vuillet L, Lajus A, Cruveiller S, Rouy Z, Mangenot S, Segurens B, Dossat C, Franck WL, Chang WS, Saunders E, Bruce D, Richardson P, Normand P, Dreyfus B, Pignol D, Stacey G, Emerich D, Vermeglio A, Medigue C, Sadowsky M.** 2007. Legumes symbioses: absence of Nod genes in photosynthetic bradyrhizobia. Science **316:**1307–1312. http://dx.doi.org/10.1126/science.1139548.
6. **Miche L, Moulin L, Chaintreuil C, Contreras-Jimenez JL, Munive-Hernandez JA, Villegas-Hernandez MD, Crozier F, Bena G.** 2010. Diversity analyses of *Aeschynomene* symbionts in Tropical Africa and Central America reveal that *nod*-independent stem nodulation is not restricted to photosynthetic bradyrhizobia. Environ. Microbiol. **12:**2152–2164. http://dx.doi.org/10.1111/j.1462-2920.2009.02090.x.
7. **Amadou C, Pascal G, Mangenot S, Glew M, Bontemps C, Capela D, Carrère S, Cruveiller S, Dossat C, Lajus A, Marchetti M, Poinsot V, Rouy Z, Servin B, Saad M, Schenowitz C, Barbe V, Batut J, Médigue C, Masson-Boivin C.** 2008. Genome sequence of the beta-rhizobium *Cupriavidus taiwanensis* and comparative genomics of rhizobia. Genome Res. **18:**1472–1483. http://dx.doi.org/10.1101/gr.076448.108.
8. **Man CX, Wang H, Chen WF, Sui XH, Wang ET, Chen WX.** 2008. Diverse rhizobia associated with soybean grown in the subtropical and tropical regions of China. Plant Soil **310:**77–87. http://dx.doi.org/10.1007/s11104-008-9631-3.
9. **Li QQ, Wang ET, Zhang YZ, Zhang YM, Tian CF, Sui XH, Chen WF, Chen WX.** 2011. Diversity and biogeography of rhizobia isolated from root nodules of *Glycine max* grown in Hebei province, China. Microb. Ecol. **61:**917–931. http://dx.doi.org/10.1007/s00248-011-9820-0.
10. **Zhang YM, Li Y, Jr, Chen WF, Wang ET, Tian CF, Li QQ, Sui XH, Chen WX.** 2011. Biodiversity and biogeography of rhizobia associated with soybean plants grown in the North China Plain. Appl. Environ. Microbiol. **77:**6331–6342. http://dx.doi.org/10.1128/AEM.00542-11.
11. **Vinuesa P, Rojas-Jimenez K, Contreras-Moreira B, Mahna SK, Prasad BN, Moe H, Selvaraju SB, Thierfelder H, Werner D.** 2008. Multilocus sequence analysis for assessment of the biogeography and evolutionary genetics of four *Bradyrhizobium* species that nodulate soybeans on the Asiatic continent. Appl. Environ. Microbiol. **74:**6987–6996. http://dx.doi.org/10.1128/AEM.00875-08.
12. **Tang J, Bromfield ESP, Rodrigue N, Cloutier S, Tambong JT.** 2012. Microevolution of symbiotic *Bradyrhizobium* populations associated with soybeans in east North America. Ecol. Evol. **2:**2943–2961. http://dx.doi.org/10.1002/ece3.404.
13. **Shiro S, Matsuura S, Saiki R, Sigua GC, Yamamoto A, Umehara Y, Hayashi M, Saeki Y.** 2013. Genetic diversity and geographical distribution of indigenous soybean-nodulating bradyrhizobia in the United States. Appl. Environ. Microbiol. **79:**3610–3618. http://dx.doi.org/10.1128/AEM.00236-13.
14. **Li QQ, Wang ET, Chang YL, Zhang YZ, Zhang YM, Sui XH, Chen WF, Chen WX.** 2011. *Ensifer sojae* sp. nov., isolated from root nodules of *Glycine max* grown in saline-alkaline soils. Int. J. Syst. Evol. Microbiol. **61:**1981–1988. http://dx.doi.org/10.1099/ijs.0.025049-0.
15. **Saldana G, Martinez-Alcantara V, Vinardell JM, Bellogin R, Ruiz-Sainz JE, Balatti PA, Saldaña G, Martinez-Alcántara V, Bellogín R, Ruíz-Sainz JE.** 2003. Genetic diversity of fast-growing rhizobia that nodulate soybean (*Glycine max* L. Merr). Arch. Microbiol. **180:**45–52. http://dx.doi.org/10.1007/s00203-003-0559-y.
16. **Camacho M, Santamaria C, Temprano E, Rodriguez-Navarro DN, Daza A, Espuny R, Bellogin R, Ollero FJ, Lyrade MCCP, Buendia-Claveria A, Zhou J, Li FD, Mateos C, Velazquez E, Vinardell JM,**

Ruiz-Sainz JE, Temprano F, de Lyra, MC. 2002. Soils of the Chinese Hubei province show a very high diversity of *Sinorhizobium fredii* strains. Syst. Appl. Microbiol. **25:**592–602. http://dx.doi.org/10.1078/07232020260517733.

17. **Suzuki K, Oguro H, Yamakawa T, Yamamoto A, Akao S, Saeki Y.** 2008. Diversity and distribution of indigenous soybean-nodulating rhizobia in the Okinawa islands, Japan. Soil Sci. Plant Nutr. **54:**237–246. http://dx.doi.org/10.1111/j.1747-0765.2007.00236.x.
18. **Keyser HH, Bohlool BB, Hu TS, Weber DF.** 1982. Fast-growing rhizobia isolated from root nodules of soybean. Science **215:**1631–1632. http://dx.doi.org/10.1126/science.215.4540.1631.
19. **Albareda M, Rodriguez-Navarro DN, Temprano FJ.** 2009. Use of *Sinorhizobium* (*Ensifer*) *fredii* for soybean inoculants in South Spain. Eur. J. Agron. **30:**205–211. http://dx.doi.org/10.1016/j.eja.2008.10.002.
20. **Rodriguez-Navarro DN, Bellogin R, Camacho M, Daza A, Medina C, Ollero FJ, Santamaria C, Ruiz-Sainz JE, Vinardell JM, Temprano FJ.** 2003. Field assessment and genetic stability of *Sinorhizobium fredii* strain SMH12 for commercial soybean inoculants. Eur. J. Agron. **19:**299–309. http://dx.doi.org/10.1016/S1161-0301(02)00076-X.
21. **Tian CF, Zhou YJ, Zhang YM, Li QQ, Zhang YZ, Li DF, Wang S, Wang J, Gilbert LB, Li YR, Chen WX.** 2012. Comparative genomics of rhizobia nodulating soybean suggests extensive recruitment of lineage-specific genes in adaptations. Proc. Natl. Acad. Sci. U. S. A. **109:**8629–8634. http://dx.doi.org/10.1073/pnas.1120436109.
22. **Tian CF, Young JPW, Wang ET, Tamimi SM, Chen WX.** 2010. Population mixing of *Rhizobium leguminosarum* bv. *viciae* nodulating *Vicia faba*: the role of recombination and lateral gene transfer. FEMS Microbiol. Ecol. **73:**563–576. http://dx.doi.org/10.1111/j.1574-6941.2010.00909.x.
23. **Silva C, Vinuesa P, Eguiarte LE, Souza V, Martínez-Romero E.** 2005. Evolutionary genetics and biogeographic structure of *Rhizobium gallicum* sensu lato, a widely distributed bacterial symbiont of diverse legumes. Mol. Ecol. **14:**4033–4050. http://dx.doi.org/10.1111/j.1365-294X.2005.02721.x.
24. **Vinuesa P, Silva C, Werner D, Martinez-Romero E.** 2005. Population genetics and phylogenetic inference in bacterial molecular systematics: the roles of migration and recombination in *Bradyrhizobium* species cohesion and delineation. Mol. Phylogenet. Evol. **34:**29–54. http://dx.doi.org/10.1016/j.ympev.2004.08.020.
25. **Krishnan HB, Natarajan SS, Kim WS.** 2011. Distinct cell surface appendages produced by *Sinorhizobium fredii* USDA257 and *S. fredii* USDA191, cultivar-specific and nonspecific symbionts of soybean. Appl. Environ. Microbiol. **77:**6240–6248. http://dx.doi.org/10.1128/AEM.05366-11.
26. **Zhang YM, Tian CF, Sui XH, Chen WF, Chen WX.** 2012. Robust markers reflecting phylogeny and taxonomy of rhizobia. PLoS One **7:**e44936. http://dx.doi.org/10.1371/journal.pone.0044936.
27. **Schmeisser C, Liesegang H, Krysciak D, Bakkou N, Le Quere A, Wollherr A, Heinemeyer I, Morgenstern B, Pommerening-Roser A, Flores M, Palacios R, Brenner S, Gottschalk G, Schmitz RA, Broughton WJ, Perret X, Strittmatter AW, Streit WR.** 2009. *Rhizobium* sp. strain NGR234 possesses a remarkable number of secretion systems. Appl. Environ. Microbiol. **75:**4035–4045. http://dx.doi.org/10.1128/AEM.00515-09.
28. **Schuldes J, Rodriguez Orbegoso M, Schmeisser C, Krishnan HB, Daniel R, Streit WR.** 2012. Complete genome sequence of the broad-host-range strain *Sinorhizobium fredii* USDA257. J. Bacteriol. **194:**4483. http://dx.doi.org/10.1128/JB.00966-12.
29. **Buendia-Claveria AM, Moussaid A, Ollero J, Vinardell JM, Torres A, Moreno J, Gil-Serrano AM, Rodriguez-Carvajal MA, Tejero-Mateo P, Peart JL, Brewin NJ, Ruiz-Sainz JE, Ollero FJ.** 2003. A *purL* mutant of *Sinorhizobium fredii* HH103 is symbiotically defective and altered in its lipopolysaccharide. Microbiology **149:**1807–1818. http://dx.doi.org/10.1099/mic.0.26099-0.
30. **Ardissone S, Kobayashi H, Kambara K, Rummel C, Noel KD, Walker GC, Broughton WJ, Deakin WJ.** 2011. Role of BacA in lipopolysaccharide synthesis, peptide transport, and nodulation by *Rhizobium* sp. strain NGR234. J. Bacteriol. **193:**2218–2228. http://dx.doi.org/10.1128/JB.01260-10.
31. **Mao C, Qiu J, Wang C, Charles TC, Sobral BW.** 2005. NodMutDB: a database for genes and mutants involved in symbiosis. Bioinformatics **21:**2927–2929. http://dx.doi.org/10.1093/bioinformatics/bti427.
32. **Lorio JC, Kim WS, Krishnan AH, Krishnan HB.** 2010. Disruption of the glycine cleavage system enables *Sinorhizobium fredii* USDA257 to form nitrogen-fixing nodules on agronomically improved North American soybean cultivars. Appl. Environ. Microbiol. **76:**4185–4193. http://dx.doi.org/10.1128/AEM.00437-10.
33. **Snoeck C, Verreth C, Hernández-Lucas I, Martínez-Romero E, Vanderleyden J.** 2003. Identification of a third sulfate activation system in *Sinorhizobium* sp. strain BR816: the CysDN sulfate activation complex. Appl. Environ. Microbiol. **69:**2006–2014. http://dx.doi.org/10.1128/AEM.69.4.2006-2014.2003.
34. **Machado D, Krishnan HB.** 2003. *nodD* alleles of *Sinorhizobium fredii* USDA191 differentially influence soybean nodulation, *nodC* expression, and production of exopolysaccharides. Curr. Microbiol. **47:**134–137. http://dx.doi.org/10.1007/s00284-002-3972-6.
35. **Jabbouri S, Fellay R, Talmont F, Kamalaprija P, Burger U, Relic B, Promé JC, Broughton WJ.** 1995. Involvement of *nodS* in *N*-methylation and *nodU* in 6-*O*-carbamoylation of *Rhizobium* sp. NGR234 Nod factors. J. Biol. Chem. **270:**22968–22973. http://dx.doi.org/10.1074/jbc.270.39.22968.
36. **de Lyra MDCP, Lopez-Baena FJ, Madinabeitia N, Vinardell JM, Espuny MD, Cubo MT, Bellogin RA, Ruiz-Sainz JE, Ollero FJ, de Lyra MDC, Espuny MDR, Belloguin RA.** 2006. Inactivation of the *Sinorhizobium fredii* HH103 *rhcJ* gene abolishes nodulation outer proteins (Nops) secretion and decreases the symbiotic capacity with soybean. Int. Microbiol. **9:**125–133.
37. **Bianco C, Defez R.** 2009. *Medicago truncatula* improves salt tolerance when nodulated by an indole-3-acetic acid-overproducing *Sinorhizobium meliloti* strain. J. Exp. Bot. **60:**3097–3107. http://dx.doi.org/10.1093/jxb/erp140.
38. **Theunis M, Kobayashi H, Broughton WJ, Prinsen E.** 2004. Flavonoids, NodD1, NodD2, and nod-box NB15 modulate expression of the *y4wEFG* locus that is required for indole-3-acetic acid synthesis in *Rhizobium* sp. strain NGR234. Mol. Plant Microbe Interact. **17:**1153–1161. http://dx.doi.org/10.1094/MPMI.2004.17.10.1153.
39. **Kot YH.** 1990. Nodule formation in soybeans by exopolysaccharide mutants of *Rhizobium fredii* USDA 191. J. Gen. Microbiol. **136:**105–113. http://dx.doi.org/10.1099/00221287-136-1-105.
40. **Staehelin C, Forsberg LS, D'Haeze W, Gao MY, Carlson RW, Xie ZP, Pellock BJ, Jones KM, Walker GC, Streit WR, Broughton WJ.** 2006. Exo-oligosaccharides of *Rhizobium* sp. strain NGR234 are required for symbiosis with various legumes. J. Bacteriol. **188:**6168–6178. http://dx.doi.org/10.1128/JB.00365-06.
41. **Jiang GQ, Krishnan AH, Kim YW, Wacek TJ, Krishnan HB.** 2001. A functional myo-inositol dehydrogenase gene is required for efficient nitrogen fixation and competitiveness of *Sinorhizobium fredii* USDA191 to nodulate soybean (*Glycine max* [L.] Merr.). J. Bacteriol. **183:**2595–2604. http://dx.doi.org/10.1128/JB.183.8.2595-2604.2001.
42. **Bardin S, Dan S, Osteras M, Finan TM.** 1996. A phosphate transport system is required for symbiotic nitrogen fixation by *Rhizobium meliloti*. J. Bacteriol. **178:**4540–4547.
43. **Han LL, Wang ET, Han TX, Liu J, Sui XH, Chen WF, Chen WX.** 2009. Unique community structure and biogeography of soybean rhizobia in the saline-alkaline soils of Xinjiang, China. Plant Soil **324:**291–305. http://dx.doi.org/10.1007/s11104-009-9956-6.
44. **Vincent JM.** 1970. A manual for the practical study of root nodule bacteria. Blackwell, Oxford, United Kingdom.
45. **Terefework Z, Kaijalainen S, Lindstrom K.** 2001. AFLP fingerprinting as a tool to study the genetic diversity of *Rhizobium galegae* isolated from *Galega orientalis* and *Galega officinalis*. J. Biotechnol. **91:**169–180. http://dx.doi.org/10.1016/S0168-1656(01)00338-8.
46. **Harrison PW, Lower RPJ, Kim NKD, Young JPW.** 2010. Introducing the bacterial "chromid": not a chromosome, not a plasmid. Trends Microbiol. **18:**141–148. http://dx.doi.org/10.1016/j.tim.2009.12.010.
47. **Thompson JD, Higgins DG, Gibson TJ.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:**4673–4680. http://dx.doi.org/10.1093/nar/22.22.4673.
48. **Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S.** 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. **28:**2731–2739. http://dx.doi.org/10.1093/molbev/msr121.
49. **Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O.** 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. **59:**307–321. http://dx.doi.org/10.1093/sysbio/syq010.

50. **Posada D, Crandall KA.** 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics **14:**817–818. http://dx.doi.org/10.1093/bioinformatics/14.9.817.
51. **Huson DH, Bryant D.** 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. **23:**254–267. http://dx.doi.org/10.1093/molbev/msj030.
52. **Didelot X, Falush D.** 2007. Inference of bacterial microevolution using multilocus sequence data. Genetics **175:**1251–1266. http://dx.doi.org/10.1534/genetics.106.063305.
53. **Gelman A, Rubin DB.** 1992. Inference from iterative simulation using multiple sequences. Stat. Sci. **7:**457–472. http://dx.doi.org/10.1214/ss/1177011136.
54. **Librado P, Rozas J.** 2009. DnaSP v5: a software for comprehensive analysis of DNA poly-morphism data. Bioinformatics **25:**1451–1452. http://dx.doi.org/10.1093/bioinformatics/btp187.
55. **Nei M.** 1987. Molecular evolutionary genetics. Columbia University Press, New York, NY.
56. **Pritchard JK, Stephens M, Donnelly P.** 2000. Inference of population structure using multilocus genotype data. Genetics **155:**945–959.
57. **Hubisz MJ, Falush D, Stephens M, Pritchard JK.** 2009. Inferring weak population structure with the assistance of sample group information. Mol. Ecol. Resour. **9:**1322–1332. http://dx.doi.org/10.1111/j.1755-0998.2009.02591.x.
58. **Jakobsson M, Rosenberg NA.** 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics **23:**1801–1806. http://dx.doi.org/10.1093/bioinformatics/btm233.
59. **Shimodaira H, Hasegawa M.** 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. **16:**1114–1116. http://dx.doi.org/10.1093/oxfordjournals.molbev.a026201.
60. **Swofford DL.** 1998. PAUP*. Phylogenetic analysis using parsimony (* and other methods), version 4. Sinauer Associates, Sunderland, MA.
61. **Hudson RR, Kaplan NL.** 1985. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics **111:**147–164.
62. **Heath L, van der Walt E, Varsani A, Martin DP.** 2006. Recombination patterns in aphthoviruses mirror those found in other picornaviruses. J. Virol. **80:**11827–11832. http://dx.doi.org/10.1128/JVI.01100-06.
63. **Guttman DS, Dykhuizen DE.** 1994. Clonal divergence in *Escherichia coli* as a result of recombination, not mutation. Science **266:**1380–1383. http://dx.doi.org/10.1126/science.7973728.
64. **Milkman R, Bridges MM.** 1990. Molecular evolution of the *Escherichia coli* chromosome. III. Clonal frames. Genetics **126:**505–517.
65. **Biondi EG, Pilli E, Giuntini E, Roumiantseva ML, Andronov EE, Onichtchouk OP, Kurchak ON, Simarov BV, Dzyubenko NI, Mengoni A, Bazzicalupo M.** 2003. Genetic relationship of *Sinorhizobium meliloti* and *Sinorhizobium medicae* strains isolated from Caucasian region. FEMS Microbiol. Lett. **220:**207–213. http://dx.doi.org/10.1016/S0378-1097(03)00098-3.
66. **Silva C, Kan FL, Martínez-Romero E.** 2007. Population genetic structure of *Sinorhizobium meliloti* and *S. medicae* isolated from nodules of *Medicago* spp. in Mexico. FEMS Microbiol. Ecol. **60:**477–489. http://dx.doi.org/10.1111/j.1574-6941.2007.00301.x.
67. **Bailly X, Olivieri I, Brunel B, Cleyet-Marel JCC, Béna G.** 2007. Horizontal gene transfer and homologous recombination drive the evolution of the nitrogen-fixing symbionts of *Medicago* species. J. Bacteriol. **189:**5223–5236. http://dx.doi.org/10.1128/JB.00105-07.
68. **Sugawara M, Epstein B, Badgley BD, Unno T, Xu L, Reese J, Gyaneshwar P, Denny R, Mudge J, Bharti AK, Farmer AD, May GD, Woodward JE, Médigue C, Vallenet D, Lajus A, Rouy Z, Martinez-Vaz B, Tiffin P, Young ND, Sadowsky MJ.** 2013. Comparative genomics of the core and accessory genomes of 48 *Sinorhizobium* strains comprising five genospecies. Genome Biol. **14:**R17. http://dx.doi.org/10.1186/gb-2013-14-2-r17.
69. **Bailly X, Giuntini E, Sexton MC, Lower RP, Harrison PW, Kumar N, Young JPW.** 2011. Population genomics of *Sinorhizobium medicae* based on low-coverage sequencing of sympatric isolates. ISME J. **5:**1722–1734. http://dx.doi.org/10.1038/ismej.2011.55.
70. **Galardini M, Pini F, Bazzicalupo M, Biondi EG, Mengoni A.** 2013. Replicon-dependent bacterial genome evolution: the case of *Sinorhizobium meliloti*. Genome Biol. Evol. **5:**542–558. http://dx.doi.org/10.1093/gbe/evt027.
71. **Epstein B, Branca A, Mudge J, Bharti AK, Briskine R, Farmer AD, Sugawara M, Young ND, Sadowsky MJ, Tiffin P.** 2012. Population genomics of the facultatively mutualistic bacteria *Sinorhizobium meliloti* and *S. medicae*. PLoS Genet. **8:**e1002868. http://dx.doi.org/10.1371/journal.pgen.1002868.
72. **Cooper JE.** 2004. Multiple responses of rhizobia to flavonoids during legume root infection. Adv. Bot. Res. **41:**1–62. http://dx.doi.org/10.1016/S0065-2296(04)41001-5.
73. **Guan SH, Gris C, Cruveiller S, Pouzet C, Tasse L, Leru A, Maillard A, Médigue C, Batut J, Masson-Boivin C, Capela D.** 2013. Experimental evolution of nodule intracellular infection in legume symbionts. ISME J. **7:**1367–1377. http://dx.doi.org/10.1038/ismej.2013.24.
74. **Marchetti M, Capela D, Glew M, Cruveiller S, Chane-Woon-Ming B, Gris C, Timmers T, Poinsot V, Gilbert LB, Heeb P, Medigue C, Batut J, Masson-Boivin C.** 2010. Experimental evolution of a plant pathogen into a legume symbiont. PLoS Biol. **8:**e1000280. http://dx.doi.org/10.1371/journal.pbio.1000280.
75. **Bianco C, Defez R.** 2010. Improvement of phosphate solubilization and *Medicago* plant yield by an indole-3-acetic acid-overproducing strain of *Sinorhizobium meliloti*. Appl. Environ. Microbiol. **76:**4626–4632. http://dx.doi.org/10.1128/AEM.02756-09.
76. **Camerini S, Senatore B, Lonardo E, Imperlini E, Bianco C, Moschetti G, Rotino GL, Campion B, Defez R.** 2008. Introduction of a novel pathway for IAA biosynthesis to rhizobia alters vetch root nodule development. Arch. Microbiol. **190:**67–77. http://dx.doi.org/10.1007/s00203-008-0365-7.
77. **Lozano L, Hernandez-Gonzalez I, Bustos P, Santamaria RI, Souza V, Young JPW, Davila G, Gonzalez V.** 2010. Evolutionary dynamics of insertion sequences in relation to the evolutionary histories of the chromosome and symbiotic plasmid genes of *Rhizobium etli* populations. Appl. Environ. Microbiol. **76:**6504–6513. http://dx.doi.org/10.1128/AEM.01001-10.