JOURNAL OF BACTERIOLOGY, July 2007, p. 5223–5236 Vol. 189, No. 14
0021-9193/07/$08.00+0 doi:10.1128/JB.00105-07

Horizontal Gene Transfer and Homologous Recombination Drive the Evolution of the Nitrogen-Fixing Symbionts of *Medicago* Species[▽]

Xavier Bailly,[1,2]* Isabelle Olivieri,[2] Brigitte Brunel,[1] Jean-Claude Cleyet-Marel,[1] and Gilles Béna[1]

Laboratoire des Symbioses Tropicales et Méditerranéennes, UMR 113 IRD-Cirad-SupAgro-UM2/USC INRA, Campus de Baillarguet, 34398 Montpellier Cedex 5, France,[1] and Institut des Sciences de l'Evolution de Montpellier, UMR 5554 CNRS-UM2, Montpellier, France[2]

Received 18 January 2007/Accepted 1 May 2007

Using nitrogen-fixing *Sinorhizobium* species that interact with *Medicago* plants as a model system, we aimed at clarifying how sex has shaped the diversity of bacteria associated with the genus *Medicago* on the interspecific and intraspecific scales. To gain insights into the diversification of these symbionts, we inferred a topology that includes the different specificity groups which interact with *Medicago* species, based on sequences of the nodulation gene cluster. Furthermore, 126 bacterial isolates were obtained from two soil samples, using *Medicago truncatula* and *Medicago laciniata* as host plants, to study the differentiation between populations of *Sinorhizobium medicae*, *Sinorhizobium meliloti* bv. meliloti, and *S. meliloti* bv. medicaginis. The former two can be associated with *M. truncatula* (among other species of *Medicago*), whereas the last organism is the specific symbiont of *M. laciniata*. These bacteria were characterized using a multilocus sequence analysis of four loci, located on the chromosome and on the two megaplasmids of *S. meliloti*. The phylogenetic results reveal that several interspecific horizontal gene transfers occurred during the diversification of *Medicago* symbionts. Within *S. meliloti*, the analyses show that *nod* genes specific to different host plants have spread to different genetic backgrounds through homologous recombination, preventing further divergence of the different ecotypes. Thus, specialization to different host plant species does not prevent the occurrence of gene flow among host-specific biovars of *S. meliloti*, whereas reproductive isolation between *S. meliloti* bv. meliloti and *S. medicae* is maintained even though these bacteria can cooccur in sympatry on the same individual host plants.

Assessing the impact of gene movements on the diversification of bacteria is one of the most intriguing issues raised by the large-scale availability of sequence data, i.e., whole genome sequences and intraspecific multilocus sequence analysis (MLSA) data sets (45). Sex (i.e., genetic exchanges) is likely to have consequences for both adaptive evolution and diversification of bacteria.

First, sex can promote adaptive evolution through various mechanisms, some of which are specific to bacteria. For instance, illegitimate recombination or horizontal gene transfer (HGT), leading to the insertion of new DNA sequences, could enhance the ecological niche of a bacterial species through the acquisition of new metabolic capabilities (21). By placing homologous adaptive mutations in different genetic backgrounds, homologous recombination could uncouple the evolution of different genomic regions (30) and thus also facilitate adaptive evolution.

Second, the addition of a new DNA fragment through HGT could promote speciation among bacterial lineages, as this would lower the rate of homologous recombination among initially promiscuous lineages (55). It has also been suggested that disruptive selection, because it promotes sequence divergence at loci undergoing such selection, might lead to overall lower rates of recombination and thus to increased divergence among lineages (8). Yet homologous recombination might be an efficient cohesion force that prevents bacterial speciation in a neutral Fisher-Wright model (24). In this context, empirical studies focusing on the interplay between ecological adaptation and sex are needed to gain a better knowledge of the processes leading to bacterial diversification.

Nitrogen-fixing bacteria that interact with leguminous plants belonging to the genus *Medicago* appear to be an ideal biological model for such studies. Although only three bacterial species have been described as the usual symbionts of *Medicago* plants, Bena et al. (3) showed that the genus *Medicago* includes at least five groups of species featuring different symbiotic specificities, as follows. (i) The sister species *Medicago laciniata* and *Medicago sauvagei* interact efficiently with a specific ecotype of *Sinorhizobium meliloti* called *S. meliloti* bv. medicaginis (56). (ii) A paraphyletic group of *Medicago* species, which includes *Medicago rigiduloides*, *Medicago noeana*, and *Medicago radiata*, interacts with another ecotype of *S. meliloti*. Most species belonging to the last emerging clade of *Medicago* species can benefit from nitrogen fixation by interacting with either (iii) *Sinorhizobium medicae* (e.g., *Medicago polymorpha*) or (iv) *S. medicae* and *S. meliloti* bv. meliloti (e.g., *Medicago truncatula*) (38). (v) Finally, *Rhizobium mongolense*, which is also described as *Rhizobium gallicum* bv. orientale (43), is the specific symbiont of *Medicago ruthenica* (53).

Since the sequencing of the genome of *S. meliloti* strain 1021 (16), considerable efforts have been made to characterize the impact of sex on the genetic structure of *S. meliloti* bv. meliloti and *S. medicae* populations (1, 49, 54). Within the *S. meliloti* bv. meliloti strain 1021 genome, the chromosome harbors mainly core genes, encoding functions related to central me-

* Corresponding author. Mailing address: Laboratoire des Symbioses Tropicales et Méditerranéennes, UMR 113 IRD-Cirad-Ensam-UM2/USC INRA, Campus de Baillarguet, 34398 Montpellier Cedex 5, France. Phone: 33 (0)4 67 59 38 01. Fax: 33 (0)4 67 59 38 02. E-mail: bailly@isem.univ-montp2.fr.

▽ Published ahead of print on 11 May 2007.

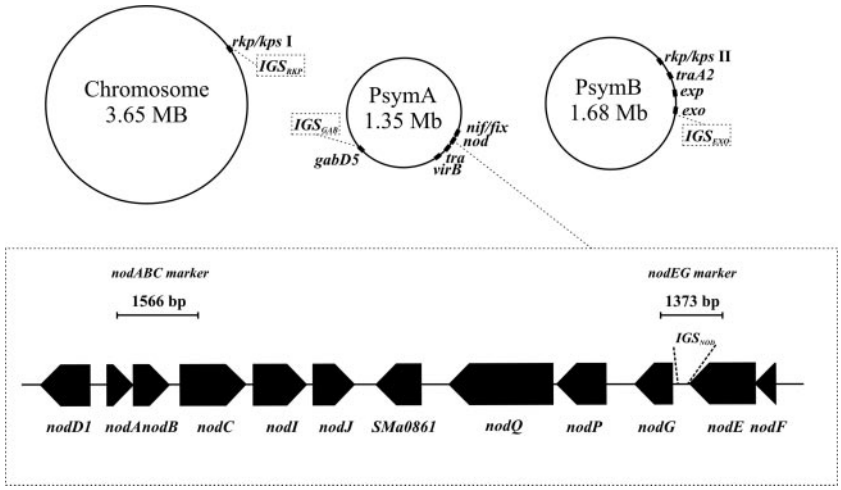

FIG. 1. Locations of the genetic markers used in this study on the genome of *Sinorhizobium meliloti* strain 1021. The four loci used to perform the MLSA study are framed by dotted lines, and their locations are indicated by dotted lines. Gene clusters located near each genetic marker are indicated by black boxes. It is noteworthy that the IGS_{NOD} marker is located near genes involved in symbiotic specificity (*nod* genes), symbiotic efficiency (*nif/fix* genes), secretion (*virB* gene), and conjugation (*tra* genes). The locations of the partial sequences of the *nodABC* and *nodEG* genes used to obtain a supertree are indicated above the schematic representation of the symbiotic gene area.

tabolism and informational processing, whereas the pSymA and pSymB megaplasmids harbor most of the accessory genes, which commonly encode supplementary metabolic pathways and symbiotic factors. Specifically, bacterial genes involved in recognition by host plants through Nod factor synthesis, such as *nod*, *noe*, and *nol* genes, are localized on pSymA. Furthermore, several genes that encode surface polysaccharidic compound pathways (lipopolysaccharide, exopolysaccharide, and capsular polysaccharide), e.g., *exo* genes, are located on pSymB (Fig. 1). Using a comparative genomic hybridization approach, Giuntini et al. (18) found that the megaplasmids of *S. meliloti*, especially pSymA, behave as hot spots of gain and loss of genes. Moreover, MLSA showed that homologous recombination involving plasmidic loci is quite common within both *S. meliloti* bv. meliloti and *S. medicae*, leading to linkage equilibrium among the three replication units (1, 49) and within megaplasmids (1). On an interspecific scale, genetic studies suggested that *S. meliloti* bv. meliloti and *S. medicae* are sexually isolated, even though they can cooccur on the same individual host plants (1, 31, 54). Nevertheless, nothing is known about the genetic differentiation among *S. meliloti* ecotypes, in particular between *S. meliloti* bv. meliloti and *S. meliloti* bv. medicaginis, which are the only two *S. meliloti* biovars formally described so far as *Medicago* symbionts (56). Populations of these two biovars can be found in sympatry, as can the two species *S. meliloti* and *S. medicae*. Barran et al. (2) described polymorphism in the *nodC* gene as the main determinant of the variation in host range between the two *S. meliloti* biovars. Such specificity could induce disruptive selective pressures as well as reproductive isolation mediated by the host plant, thus promoting the coexistence of conspecific ecotypes at the community level. This process might then lead to the divergence of ecotypes for genes not involved in host specificity (i.e., speciation) due to selective sweeps acting on each ecotype and/or sexual isolation. Conversely, homogenizing selection acting on the core genome (i.e., purifying selection and global selective sweeps) and frequent recombination events might prevent strains specialized from different *Medicago* species, e.g., *M. laciniata* or *M. truncatula*, from diverging in genes not involved in specificity and thus from undergoing speciation (26).

In this context, the present study aims at understanding how sex, mediated by HGT or homologous recombination, has influenced the diversification of bacteria able to interact with *Medicago* species. To fulfill this goal, we first inferred a phylogeny from partial sequences of the *nodABC-nodEG* gene clusters, including symbionts of each specificity group within the genus *Medicago* described up to now. We then characterized two communities of symbiotic bacteria, including isolates belonging to *S. medicae*, *S. meliloti* bv. meliloti, and *S. meliloti* bv. medicaginis, by MLSA, an approach derived from multilocus sequence typing (25). Using MLSA, we first wanted to test the hypothesis that differentiation among populations belonging to the same species of *Sinorhizobium* was less prevalent than differentiation among sympatric populations of different species. We further wanted to test the hypothesis that disruptive selection, likely induced by host specificity, would induce overall divergence among ecotypes, possibly leading to further reproductive isolation and speciation.

MATERIALS AND METHODS

Isolation and culture of bacteria. Two samples of topsoil were obtained from rangelands located near two cities in Tunisia which are included in a semiarid bioclimatic area, namely, Enfidha (longitude, 10°23′E; latitude, 36°07′N) and Hadjeb (longitude, 9°33′E; latitude, 35°24′N). These sampling sites fall within the natural geographic ranges of both *M. laciniata* and *M. truncatula* (44). Aliquots of each soil sample were put into contact with two plant species, including (i) 40 individuals of *M. truncatula* (Jemalong A17 line) and (ii) 20 individuals of *M. laciniata* (the bulk of Tunisian genotypes were provided by the Australian *Medicago* Genetic Resource Centre located at Adelaide, Australia). The difference in genetic heterogeneity between the two sets of individuals should not affect the results, as in a previous experiment, the bacterial diversity sampled from a single genotype of *M. truncatula* was not significantly different from that sampled from 20 different genotypes of this species (1). Gibson tubes were previously filled 9/10 with vermiculite and 25 ml of nitrogen-free plant nutrient solution. Tubes were autoclaved and then filled with a soil aliquot. Seeds were

TABLE 1. Primers used in this study

Experiment	Genome region and location (bp)[a]	Primer[b]	Prime sequence (5′→3′)	Melting temp (°C)	PCR product size (bp)
MLSA	IGS_{RKP} (*rkpA-rkpU*)	RKP1	AGGCATGCACGCCCTATGA	58.8	599
		RKP2	GCCATCGACATCTACAATATCAA	57.1	599
	IGS_{NOD} (*nodE-nodG*)	NOD1	CAGTTCTGGCATTCAAGC	53.7	579
		NOD2	CCCCTCCTATGGCTCCTGAT	61.4	579
	IGS_{GAB} (*gabD5*-Sma1850)	GAB1	CATGACCAAAGACCGCTTCC	59.4	658
		GAB2	GCATGATCGGCCTCAACAC	58.8	658
	IGS_{EXO} (*exoP-thiD*)	EXOmeliloti1	CAACAAGACGGATATGAACGAA	56.5	494
		EXOmeliloti2	GTGGTGGAAGGATTGACTGC	59.4	494
		EXOmedicae1	CATGAACGAGCTGGGCAAAT	57.3	250
		EXOmedicae2	CTGGTCGAAGCGGCAAAA	56.0	250
nodABC supertree construction	pSymA (481186)	$XNA2^F$	CGCCTTTGGGACAGTTCG	58.2	
	pSymA (480970)	$XNA3^F$	CGGATCGGAGCTATGAAGCA	59.4	
	pSymA (480469)	$XNA4^R$	AGTCCAGCACTGCATCAACAAT	58.4	
	pSymA (480258)	$XNA5^R$	AGCGGCGGTGCTGGTTG	60.0	
	pSymA (480016)	$XNA6^R$	TTCCGAGAACCATCATCAACGAC	60.6	
	pSymA (479704)	$XNA8^F$	CGGCACAGTCTCGCTTCG	60.5	
	pSymA (479686)	$XNA7^R$	CCGAAGCGAGACTGTGCC	60.5	
	pSymA (479535)	$XNA9^R$	GCCTGCCTTCAACATGAGAAT	57.9	
nodEG supertree construction	pSymA (470195)	$XNE3^F$	AAATGCCCTTCGGTTCGG	56.0	
	pSymA (470427)	$XNE4^F$	GCGGTAGACCAGATCAAGTGC	61.8	
	pSymA (470576)	$XNE5^F$	CGTGGTACTGGGCGAGGGT	63.1	
	pSymA (470874)	$XNE6^F$	TCCATATCTTCCACCAAGTCCA	58.4	
	pSymA (471038)	$XNE7^F$	GCGTGAGCGTAAGGTGCG	60.5	
	pSymA (471725)	$XNE8^R$	CCGTGCAGTCCGACGAT	57.6	
	pSymA (471878)	$XNE9^R$	GCATTGTTGACCAGGATGTCG	59.8	

[a] Genome locations refer to the genome of *S. meliloti* strain 1021.
[b] F and R, forward and reverse primers, respectively.

surface sterilized with 5% (wt/vol) calcium hypochlorite for 5 min and rinsed with sterile water. Seeds were germinated for 72 h on 1% (wt/vol) agar medium. Each plant was transferred individually into a Gibson tube, which was placed in a growth chamber at 22°C during the day and at 18°C at night, with a 16-hour photoperiod and 50 to 60% relative humidity. After 2 months, all nodules from each plant were harvested. One or two nodules per plant were sterilized using 1% (wt/vol) hypochlorite for 3 minutes and then rinsed three times with sterile water. The isolation of one bacterial strain per nodule was performed using three successive subcultures of an isolated colony on yeast extract mannitol (YEM) agar (57), starting from crushed nodules. Bacterial strains were preserved at −80°C in YEM medium supplemented with 20% (vol/vol) glycerol.

Each bacterial isolate was then grown in 20 ml YEM medium. We also grew the following in 20 ml YEM medium: (i) *S. medicae* strain STM 1605 and *S. meliloti* bv. meliloti strain STM 1604 (1); (ii) three *S. meliloti* strains which had been described as interacting specifically with *M. noeana*, *M. radiata*, and *M. rigiduloides*, namely, USDA 1613, USDA 1614, and USDA 1623 (3); and (iii) *Rhizobium mongolense* strain USDA 1844, which fixes nitrogen with *M. ruthenica* (53). We used these strains to obtain a *nod* gene phylogeny illustrating the diversity of the specificity groups described so far as being associated with *Medicago* species.

DNA extraction and sequencing. (i) Experimental procedures. One milliliter of liquid culture was washed twice in an Eppendorf microcentrifuge tube by centrifugation (15,000 × *g*, 4 min), and the pellet was resuspended in 750 μl of sterile water. One hundred microliters of this solution was incubated for 2 hours with 20 μl of 1-mg/ml proteinase K and 100 μl of Tris-HCl (10 mM, pH 8.3). After boiling, this mixture was used as a DNA matrix. DNA amplification was performed using a Perkin-Elmer 2400 thermocycler with 25-μl volume reaction mixtures containing 1 μl of DNA matrix, a 200 μM concentration of each deoxynucleoside triphosphate, a 0.8 μM concentration of each primer, 1.5 mM of $MgCl_2$, 1× buffer supplied by the *Taq* polymerase manufacturer, and 1.25 U of Invitrogen *Taq* polymerase. We used a touchdown program including an initial denaturation stage (96°C, 4 min); 20 cycles of denaturation (96°C, 30 s), annealing (annealing temperature decreased steadily from 60°C to 50°C in 20 cycles, 30 s), and elongation (72°C, 1 min); and 20 further cycles of denaturation (96°C, 30 s), annealing (50°C, 30 s), and elongation (72°C, 1 min). PCR products were cut from a 1% (wt/vol) agarose electrophoresis gel and purified with a QIAquick gel extraction kit (QIAGEN). Sequencing reactions were performed on one strand by using a DYEnamic ET Terminator kit (Amersham Bioscience) and were analyzed on an Amersham Bioscience Megabace 1000 DNA sequencer.

(ii) Phylogenetic markers defined within the *nod* gene cluster of the symbionts of *Medicago* species. We aimed at obtaining the following two sequence data sets for the *nod* gene region of rhizobia that interact with *Medicago* plants (Fig. 1): (i) partial sequences of the *nodABC* gene cluster, which is ubiquitous to the genomes of all rhizobia; and (ii) partial sequences of the *nodEG* gene cluster, which seems to be specific to *Medicago* symbionts. The wide distribution of the *nodABC* gene cluster allowed us to obtain a phylogeny rooted by an appropriate outgroup that includes a "core collection" of the diversity of *nod* genes among *Rhizobiaceae.* Using the *nodEG* data set allowed us to perform an analysis that takes into account all the diversity we observed within the symbiotic cluster. Primers were designed based on homologous regions conserved among rhizobia, obtained from the GenBank database (Table 1).

(iii) MLSA scheme. In addition to the phylogenetic approach described above, we characterized symbiotic bacteria sampled from *M. truncatula* and *M. laciniata* by sequencing four loci. Previous studies have shown a significant and very strong linkage disequilibrium among chromosomal markers in *S. meliloti* bv. meliloti (1, 49). In the present study, we therefore analyzed only one locus on the chromosome rather than three loci as in our previous study. More precisely, we investigated genetic polymorphisms of one chromosomal sequence (IGS_{RKP}), one pSymB sequence (IGS_{EXO}), and two pSymA sequences (IGS_{NOD} and IGS_{GAB}) (Fig. 1). Sequences of PCR/sequencing primers are provided in Table 1.

Data analyses based on the *nod* gene clusters of the symbionts of both *Medicago* plants and other *Rhizobiaceae*. (i) Alignment of sequence data sets. Nucleotidic sequences of the *nodABC-nodEG* gene cluster were obtained from *S. medicae* strain STM 1605; *S. meliloti* bv. meliloti strain STM 1604; *S. meliloti* bv. medicaginis isolates STM 2835 and STM 2836, from this study; *S. meliloti* strains USDA 1613, USDA 1614, and USDA 1623, which interact efficiently with *Medicago rigiduloides*, *Medicago noeana*, or *Medicago radiata* (3); and *Rhizobium mongolense* strain USDA 1844. These sequences were aligned using Clustal X 1.83 (51), together with both sequences available for the marker IGS_{NOD} and with those of *S. meliloti* bv. meliloti strain 1021 (*nodABC-nodEG*; GenBank accession number [AN], NC_003037), strain 042B (*nodABC*; AN, AF038577), strain CC2093 (*nodA*; AN, AJ300226), strain CC2017 (*nodA*; AN, AJ300222), strain 102F28 (*nodA*; AN, AJ300221), and strain L5-30 (*nodA*; AN, AJ300220); *S. meliloti* bv. medicaginis strain USDA 1170 (*nodABC*; AN, AF522456); *S.*

medicae strain A321 (*nodA*, *nodB*, and *nodC*; AN, AJ300224, AY929593, and EF428921, respectively) and strain M3 (*nodA*; AN, AJ300225); *Rhizobium galegae* strain HAMBI 1174 (*nodABC*; AN, X87578) and strain HAMBI 1207 (*nodA*; AN, AJ300240); *Rhizobium tropici* strain CFN 299 (*nodABC*; AN, X98514); *Rhizobium leguminosarum* bv. viciae strain 3841 (*nodABC*; AN, AM236084) and strain 248 (*nodABC*; AN, Y00548); *R. leguminosarum* bv. trifolii strain TA1 (*nodA*; AN, AY904443); *R. mongolense* strain USDA 1832 (*nodB*; AN, AY929577), strain USDA 1834 (*nodB*; AN, AY929578), strain USDA 1836 (*nodB*; AN, AY929579), strain USDA 1849 (*nodB*; AN, AY929580), strain USDA 1877 (*nodB*; AN, AY929581), strain USDA 1890 (*nodB*; AN, AY929582), strain USDA 1904 (*nodB*; AN, AY929583), and strain USDA 1929 (*nodB*; AN, AY929584); *Rhizobium yanglingense* strain SH22623 (*nodB*; AN, AY929585); *Rhizobium giardini* strain H152 (*nodA* and *nodC*; AN, AJ300238 and AF217267), strain Ro84 (*nodA*; AN, AJ300239), and strain H251 (*nodC*; AN, AF217264); *R. gallicum* strain PhD12 (*nodA*, *nodB*, and *nodC*; AN, AJ300237, AY929556, and AF217265, respectively), strain R602sp (*nodA*, *nodB*, and *nodC*; AN, AJ300236, AF29022, and AF217266, respectively), and strain FL27 (*nodB* and *nodC*; AN, AF29021 and AF217270); *Rhizobium etli* strain CFN42 (*nodABC*; AN, U80928) and strain GR-12 (*nodB*; AN, AY929587); *Sinorhizobium* strain NGR234 (*nodABC*; AN, U00090); *Sinorhizobium terangae* strain ORS1073 (*nodA*; AN, AJ300229) and strain ORS604 (*nodA*; AN, AJ300228); and *Sinorhizobium saheli* strain ORS611 (*nodA*; AN, AJ300227). Coding sequences were aligned based on amino acid translations, while intergenic spacers (IGS) were aligned directly. Alignments were edited with BioEdit 5.0.9 (19).

(ii) Detection of recombination. We searched for traces of recombination in a concatenated data set including full sequences of *nodABC* and *nodEG* markers belonging to *Medicago* symbionts, using three methods implemented in RDP2 (i.e., MaxChi, Chimaera, and Geneconv), as these methods were the most accurate for detecting recombination events (29). The first two methods use a sliding window to detect significant discrepancies in the segregation of polymorphisms on either side of each sequence site. These are based on the computation of χ^2 distributed statistics obtained from comparisons of pairs and triplets of sequences, respectively. Geneconv is an extension of Sawyer's method (40). It compares sequence pairs and scores unusually long series of polymorphic sites for which two sequences share a strong similarity. The null hypothesis of a lack of gene conversion events (i.e., homologous recombination) is tested through the randomization of the location of aligned sites. A *P* value is given as the proportion of permuted alignments which result in a greater score than the original alignment. For all these methods, insertion-deletion blocs were analyzed as single polymorphisms, we performed 1,000 permutation steps, and we used a *P* value cutoff of 0.05, using a Bonferroni correction for multiple tests (i.e., tests performed on several sequence sets and several sequence sites). We also searched consensus daughter sequences and consensus recombination breakpoints. For MaxChi and Chimaera, we used a variable window size of 20 polymorphic sites and 10% of variable sites, respectively. For Geneconv, similarity scores were computed using a gscale (i.e., a parameter which scales the impact of mismatches on similarity scores) set to 1.

(iii) Inference of a supertree. Taking into account the heterogeneity of our data set, we applied a supertree approach (see reference 5 for a review) to combine the information available from the different strains. The analysis of a concatenated data set in a likelihood framework is the best alternative for obtaining an accurate supertree (10). We therefore inferred a phylogeny from the alignments obtained from the *nodABC-nodEG* sequences by using MrBayes 3.1 (20). We used a Bayesian approach because its implementation in MrBayes 3.1 allows the use of mixed models for partitioned data analyses. Sequence evolution was modeled as follows: (i) each IGS or gene was considered a single partition; (ii) we assumed a different model for each partition, with parameters of the different models being estimated independently; (iii) the rate of evolution was allowed to vary among partitions; (iv) a GTR + I + G (general time reversible + invariant + gamma) model was used for IGS; (v) for coding sequences, the analysis was performed based on a codon model that uses a GTR transition probability matrix to model single-nucleotide evolution and in which we assumed variation in selection across sites (i.e., NY98 option). Two independent Metropolis-coupled Markov chain Monte Carlo searches were performed, using three chains with a temperature parameter set to 0.1, starting from a random tree and lasting 2×10^5 generations. Evidence for convergence of the two searches was obtained by examining the correlation between the posterior probabilities of individual clades and by observing the distribution of the log likelihood values over generations. We sampled a tree every 100th generation after the 1.5×10^5th generation according to convergence criteria. The topologies obtained during the two searches were pooled to compute posterior probabilities of clades and thus to assess their robustness.

(iv) Comparison of divergence and selective patterns of *nod* and housekeeping genes at the interspecific level. To study the impact of selective pressures on the pattern of genetic divergence between *S. medicae* and *S. meliloti* bv. meliloti at *nod* genes, we used DNAsp 4.0 (39) to compute both the pairwise sequence divergence, i.e., $D(x,y)$ (the number of substitutions per site, using JC69 correction), and the pairwise ratio of K_a (the number of nonsynonymous substitutions per nonsynonymous site) to K_s (the number of synonymous substitutions per synonymous site) between DNA sequences belonging to *S. medicae* and *S. meliloti* bv. meliloti. First, these measures were obtained for the symbiotic genes *nodA* and *nodB* but were also obtained from partial sequences of *nodC* and *nodE* genes. Computations were performed using sequences of strains STM 1604 and STM 1605, except for *nodA* analysis, which was performed on the *S. meliloti* 1021 sequence and the *S. medicae* A321 sequence in order to take into account the whole gene sequence. Second, the mean $D(x,y)$ value and the mean K_a/K_s ratio were computed between sets of alleles belonging to *S. meliloti* bv. meliloti and to *S. medicae*, respectively, for the partial sequences of housekeeping genes. With this aim, we used the interspecific MLSA data set published by Van Berkum et al. (54), focusing on seven genes on the genome of *S. meliloti* bv. meliloti, namely, *sucA*, *glnD*, *recA*, *asd*, *gap*, *zwf*, and *ordL2*. The allelic richness of these loci ranges from 1 to 6 alleles for *S. medicae* and from 6 to 10 alleles for *S. meliloti* bv. meliloti. A comparison of the $D(x,y)$ values and the K_a/K_s ratios obtained from the set of symbiotic genes and from the set of housekeeping genes was conducted using a Mann-Whitney U test.

MLSA of *S. meliloti* and *S. medicae* polymorphism. (i) Sequence alignment, phylogenetic inferences, and detection of recombination. Nucleotide sequence alignments were performed using Clustal X 1.83 and manually edited with BioEdit 5.0.9. We built a maximum likelihood phylogeny for each molecular marker, using PAUP* 4b10, applying models of sequence evolution determined by the likelihood ratio test implemented in Modeltest 3.7 and performing heuristic searches by the TBR algorithm, starting from the neighbor-joining tree. The four phylogenies were used in three ways. To assess the robustness of bipartitions for each data set, 100 bootstrap pseudoreplicates were assessed by maximum likelihood as indicated above.

First, pairwise distance matrices were computed for the four loci according to the obtained topologies with PAUP* 4b10. These matrices were used to compute the following parameters for each locus: π_d, the mean pairwise divergence between isolates that belong to two different groups among *S. medicae*, *S. meliloti* bv. meliloti, and *S. meliloti* bv. medicaginis; and π_s, the mean pairwise divergence within the two same groups.

Second, we investigated the congruency between sequence evolution at each locus and symbiotic abilities of bacterial strains. Using Shimodaira-Hasegawa tests (41), we compared for each locus the likelihood of the data given the best tree to the likelihood of the data based on assuming the best topology obtained by enforcing the monophyly of *S. medicae*, *S. meliloti* bv. meliloti, and *S. meliloti* bv. medicaginis. Constrained trees and congruence test significances were inferred using the relevant options implemented in PAUP* 4b10, based on the evolutionary models and search parameters described above.

Third, Shimodaira-Hasegawa tests were performed to investigate likelihood differences among sequence data sets according to the topologies inferred from each of the different markers. These analyses were performed to test the null hypothesis of congruence of the topology obtained for each marker and therefore to infer whether homologous recombination might have uncoupled the evolution of the markers we used. To reveal whether significant incongruence between topologies could be explained by recombination events involving both *S. meliloti* biovars, we concatenated sequences of the four loci to screen *S. meliloti* MLSA data sets, using methods implemented in RDP2, with the following ordering: IGS_{RKP}, IGS_{EXO}, IGS_{GAB}, and IGS_{NOD}. Assuming that all these IGS are inherited vertically, they would feature similar polymorphism patterns. Conversely, an event of homologous recombination would induce a significantly different polymorphism pattern for the transferred locus compared to those for the other markers. Pairs or triplets of concatenated sequences were thus scanned with MaxChi, Chimaera, and Geneconv, using the settings described above.

(ii) Population structure. Sequences that differed from others by at least one nucleotide position were considered different alleles. The diversity of each bacterial group (i.e., *S. medicae*, *S. meliloti* bv. meliloti, and *S. meliloti* bv. medicaginis) was assessed using the summary statistics allelic richness (R) and Nei diversity (H) (33) indices, which were computed for each of the four MLSA data sets. In order to cope with differences in sample size, these statistics were obtained for 1,000 pseudoreplicates of the original data set by bootstrapping $N_{\min}$ bacterial genotypes, where $N_{\min}$ is the minimum population size observed. Expected allelic richness for each genetic group and 95% confidence intervals were estimated for the $N_{\min}$ observed among groups. The contribution of each symbiotic group to the allelic richness of pairs of ecotypes at each locus was also computed. These

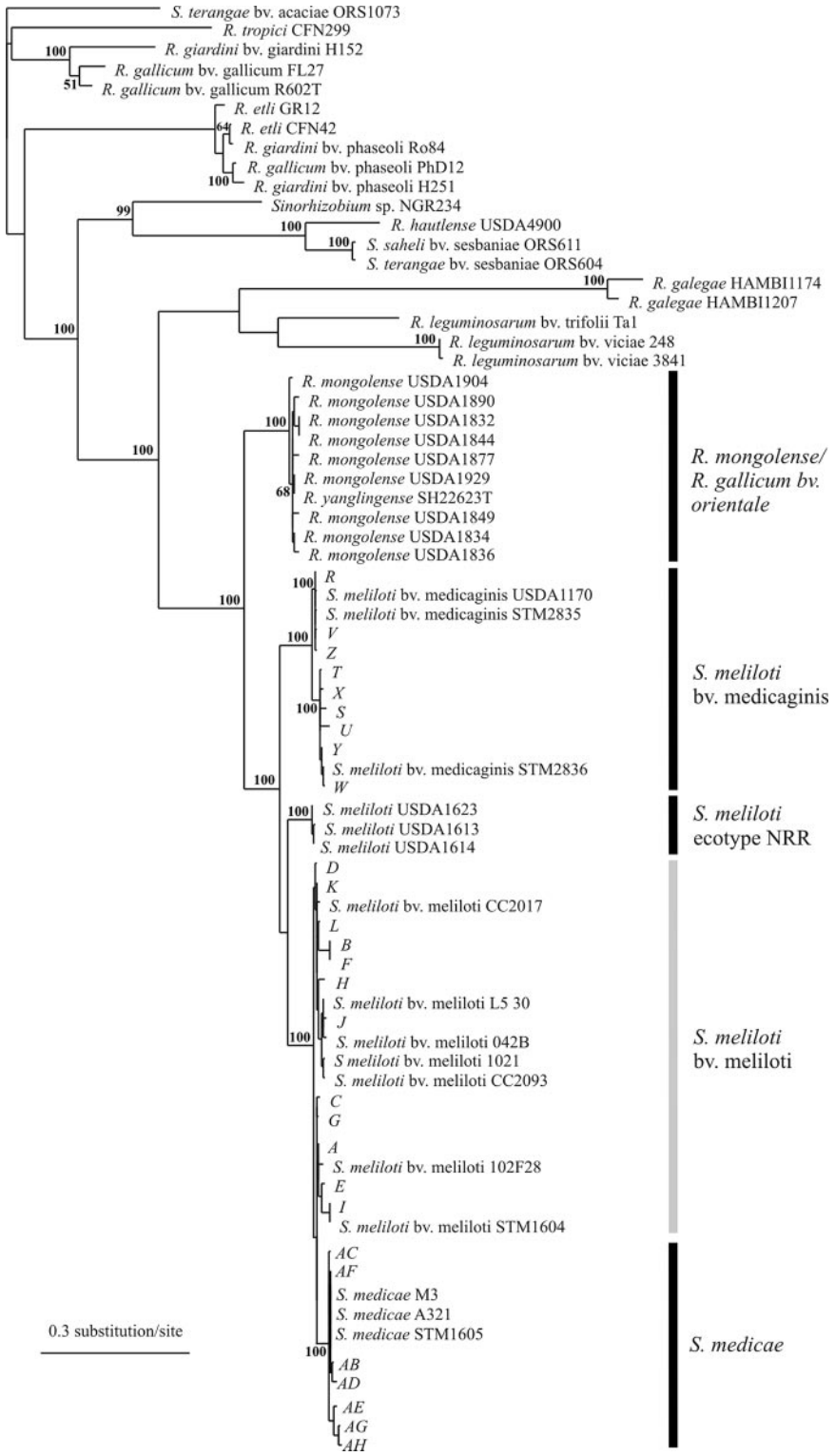

FIG. 2. Bayesian supertree obtained from *nodABC-nodEG* gene cluster. Bipartition support was assessed using the posterior probabilities of the different clades. Letters refer to the allelic diversity described for the IGS_{NOD} marker in both this study and that of Bailly et al. (1). The topology indicates that symbionts of the genus *Medicago* cluster in a monophyletic group. Within this group, four clades, which are indicated by black lines, include the following: (i) *R. mongolense*/*R. gallicum* bv. orientale isolates; (ii) *S. meliloti* bv. medicaginis isolates; (iii) isolates which cluster in a group called *S. meliloti* ecotype NRR, i.e., bacteria associated with *Medicago noeana*, *M. rigiduloides*, and *M. radiata*; and (iv) *S. medicae* isolates. Conversely, *S. meliloti* bv. meliloti, which is indicated by a gray line, is paraphyletic due to the position of the *S. medicae* clade.

TABLE 2. Distributions of overall, nonsynonymous, and synonymous divergence between *S. medicae* and *S. meliloti* bv. meliloti for a set of housekeeping genes and a set of symbiotic genes

Gene	Length (bp)	No. of fixed mutations	$D(x,y)$	K_a	K_s	K_a/K_s
sucA	432	15	0.0487	0.00356	0.2227	0.0159
glnD	311	42	0.1607	0.02291	0.8979	0.0255
recA	329	26	0.1027	0.01636	0.4095	0.0399
asd	465	32	0.0930	0.02521	0.3388	0.0744
gap	416	19	0.0626	0.01693	0.2239	0.0756
zwf	453	39	0.1132	0.03453	0.4359	0.0792
ordL2	433	38	0.1132	0.04859	0.3309	0.1468
Average for housekeeping genes			0.0992	0.02401	0.4085	0.0653
nodE	672	9	0.0135	0.0020	0.0476	0.0420
nodA	591	15	0.0258	0.0069	0.0835	0.0826
nodB	713	18	0.0256	0.0095	0.0750	0.1266
nodC	654	21	0.0328	0.0125	0.0939	0.1331
Average for *nod* genes			0.0215	0.00773	0.0750	0.0961

measures were obtained using a program written in Turbo-Pascal. Differentiation among bacterial populations/taxa was analyzed using Wright's F_{ST} values among populations according to the Weir and Cockerham procedure (59), and exact tests of genotypic differentiation (36) were performed using Genepop 3.3, updated from the work of Raymond and Rousset (37).

Nucleotide sequence accession numbers. All sequences have been deposited in the GenBank database under the following accession numbers: IGS_{RKP}, DQ405642 to DQ405767; IGS_{NOD}, DQ405467 to DQ405592; IGS_{GAB}, DQ406393 to DQ406518; IGS_{EXO}, DQ406104 to DQ406229; and *nodABC-nodEG*, DQ406568 to DQ406583.

RESULTS

Data analyses based on the *nod* gene cluster of the symbionts of both *Medicago* plants and other *Rhizobiaceae*. (i) Phylogenetic inferences. Our first aim was to obtain a phylogeny including *Medicago* symbionts based on *nod* genes. To ensure that recombination did not blur the phylogenetic signal within the *nod* gene cluster of *Medicago* symbionts, we analyzed *nodABC* and *nodEG* data sets by using three methods (i.e., MaxChi, Chimaera, and Geneconv). We did not infer clues of recombination within the *nodABC* and *nodEG* concatenated sequence data set (data not shown). Furthermore, the supertree inferred from both the *nodABC* and *nodEG* gene clusters supported the hypothesis that *R. leguminosarum* bv. viciae, *R. leguminosarum* bv. trifolii, *R. galegae*, and *Medicago* symbionts cluster in a clade (Fig. 2). Within this group, *nod* regions of symbionts associated with *Medicago* clustered in a clade. *Rhizobium mongolense* appeared to be the first emerging branch among the *Medicago* symbiont, *S. meliloti*, and *S. medicae* sequences grouping in a single clade. The *nod* gene regions of *S. meliloti* isolates, including the three ecotypes defined within this species, formed a paraphyletic group because the clade of *S. medicae* strains is nested within *S. meliloti* bv. meliloti. The supertree methodology proposed by Daubin et al. (11) also revealed the four clades described above, with these groups featuring substantial bootstrap support (data not shown). Finally, these clades were also found in the individual trees that could be obtained from the sequences of *nodA*, *nodB*, *nodC*, *nodE*, *nodG*, and the IGS of this genomic area. The most noteworthy discrepancy among the trees obtained from the different data partitions is the monophyly of *S. medicae* and *S. meliloti* bv. meliloti, which is strongly supported by *nodEG* markers.

(ii) Comparative analysis between *nod* genes and housekeeping genes. To assess whether the branching of *nodABC-nodEG* sequences of *S. medicae* within *S. meliloti* bv. meliloti involved a particularly strong purifying selective pressure, we analyzed the divergence among these two groups for *nod* genes and housekeeping genes. The values of the $D(x,y)$ and K_a/K_s statistics obtained for the different genes we studied are reported in Table 2. The Mann-Whitney U test we performed on $D(x,y)$ values indicated a significantly higher accumulation of mutations since the divergence of *S. meliloti* bv. meliloti and *S. medicae* for housekeeping genes than for *nod* genes ($P < 0.005$). It is noteworthy that the measures obtained were fairly consistent with the results presented in Fig. 3, which are described below. Conversely, although higher on average, the K_a/K_s ratio for symbiotic genes was not significantly different from the K_a/K_s ratio for housekeeping genes ($P > 0.05$).

MLSA of *S. meliloti* and *S. medicae* polymorphism. (i) Sampling experiment. To investigate the genetic structure of sympatric populations of *Sinorhizobium* associated with *Medicago* plants, we sampled Tunisian bacterial isolates from root nodules of *Medicago truncatula* and *Medicago laciniata*. One bacterial isolate was obtained from each of the 120 plants, except for six *M. truncatula* plants for which two isolates per plant were used in the experiments. The bacterial sampling included a total of 126 isolates (Table 3). With regards to symbiotic efficiency, both *M. truncatula* and *M. laciniata* were able to fix nitrogen with these bacteria (green foliage, pink nodules). We obtained 504 sequences from the four loci, allowing us to define 126 complete genotypes (Table 3). The 126 isolates clustered into two main genetic groups based on sequence homology. These groups were identified as *S. meliloti* and *S. medicae* after alignments with available sequence data sets. *Sinorhizobium meliloti* isolates sampled from *M. laciniata* were assigned to *S. meliloti* bv. medicaginis based on both their symbiotic abilities and their specific IGS_{NOD} alleles. Overall, the data set for this study included 20 *S. medicae* isolates, 66 *S. meliloti* bv. meliloti isolates, and 40 *S. meliloti* bv. medicaginis isolates. Interestingly, we recovered *S. medicae* sequences only from the soil sample obtained at Enfidha. Models of sequence evolution selected by the likelihood ratio test were HKY85 models for IGS_{NOD}, IGS_{GAB}, and IGS_{EXO} data sets and JC69 for

TABLE 3. Origins and genotypes of *S. medicae* and *S. meliloti* isolates

Location	Taxon/host plant	Isolate	Allele at locus[a]			
			IGS_{RKP}	IGS_{EXO}	IGS_{GAB}	IGS_{NOD}
Enfidha	*S. medicae*/*M. truncatula*	STM 2778	K	P	Y	AH
	S. medicae/*M. truncatula*	STM 2779	K	Q	V	AE
	S. medicae/*M. truncatula*	STM 2780	K	N	Y	AE
	S. medicae/*M. truncatula*	STM 2781	K	P	W	AF
	S. medicae/*M. truncatula*	STM 2782	K	P	T	AF
	S. medicae/*M. truncatula*	STM 2783	K	L	X	AF
	S. medicae/*M. truncatula*	STM 2784	K	P	U	AE
	S. medicae/*M. truncatula*	STM 2785	K	M	T	AF
	S. medicae/*M. truncatula*	STM 2786	K	N	U	AF
	S. medicae/*M. truncatula*	STM 2787	K	P	Y	AH
	S. medicae/*M. truncatula*	STM 2788	K	R	U	AF
	S. medicae/*M. truncatula*	STM 2789	K	P	T	AE
	S. medicae/*M. truncatula*	STM 2790	K	K	Y	AE
	S. medicae/*M. truncatula*	STM 2791	K	Q	Y	AE
	S. medicae/*M. truncatula*	STM 2792	K	K	T	AE
	S. medicae/*M. truncatula*	STM 2793	K	K	U	AE
	S. medicae/*M. truncatula*	STM 2794	K	N	Y	AE
	S. medicae/*M. truncatula*	STM 2795	K	K	U	AF
	S. medicae/*M. truncatula*	STM 2796	K	Q	U	AE
	S. medicae/*M. truncatula*	STM 2797	K	K	U	AE
	S. meliloti bv. meliloti/*M. truncatula*	STM 2798	F	B	E	K
	S. meliloti bv. meliloti/*M. truncatula*	STM 2799	C	B	C	A
	S. meliloti bv. meliloti/*M. truncatula*	STM 2800	E	D	G	K
	S. meliloti bv. meliloti/*M. truncatula*	STM 2801	A	D	J	C
	S. meliloti bv. meliloti/*M. truncatula*	STM 2802	C	B	G	A
	S. meliloti bv. meliloti/*M. truncatula*	STM 2803	C	C	F	K
	S. meliloti bv. meliloti/*M. truncatula*	STM 2804	C	B	O	K
	S. meliloti bv. meliloti/*M. truncatula*	STM 2805	C	B	G	K
	S. meliloti bv. meliloti/*M. truncatula*	STM 2806	C	B	O	K
	S. meliloti bv. meliloti/*M. truncatula*	STM 2807	A	D	C	K
	S. meliloti bv. meliloti/*M. truncatula*	STM 2808	A	D	E	K
	S. meliloti bv. meliloti/*M. truncatula*	STM 2809	C	B	K	K
	S. meliloti bv. meliloti/*M. truncatula*	STM 2810	C	B	E	K
	S. meliloti bv. meliloti/*M. truncatula*	STM 2811	C	B	L	G
	S. meliloti bv. meliloti/*M. truncatula*	STM 2812	C	B	C	A
	S. meliloti bv. meliloti/*M. truncatula*	STM 2813	C	B	O	K
	S. meliloti bv. meliloti/*M. truncatula*	STM 2814	C	B	C	A
	S. meliloti bv. meliloti/*M. truncatula*	STM 2815	C	B	C	A
	S. meliloti bv. meliloti/*M. truncatula*	STM 2816	C	B	C	A
	S. meliloti bv. meliloti/*M. truncatula*	STM 2817	C	B	O	K
	S. meliloti bv. meliloti/*M. truncatula*	STM 2818	A	B	C	K
	S. meliloti bv. meliloti/*M. truncatula*	STM 2819	C	B	O	K
	S. meliloti bv. meliloti/*M. truncatula*	STM 2820	C	B	M	K
	S. meliloti bv. meliloti/*M. truncatula*	STM 2821	E	D	C	K
	S. meliloti bv. medicaginis/*M. laciniata*	STM 2822	C	B	E	Y
	S. meliloti bv. medicaginis/*M. laciniata*	STM 2823	C	B	E	Y
	S. meliloti bv. medicaginis/*M. laciniata*	STM 2824	C	B	O	R
	S. meliloti bv. medicaginis/*M. laciniata*	STM 2825	B	B	F	S
	S. meliloti bv. medicaginis/*M. laciniata*	STM 2826	C	B	E	Z
	S. meliloti bv. medicaginis/*M. laciniata*	STM 2827	E	D	B	X
	S. meliloti bv. medicaginis/*M. laciniata*	STM 2828	C	B	O	X
	S. meliloti bv. medicaginis/*M. laciniata*	STM 2829	C	B	O	W
	S. meliloti bv. medicaginis/*M. laciniata*	STM 2830	C	B	E	Z
	S. meliloti bv. medicaginis/*M. laciniata*	STM 2831	C	B	C	Z
	S. meliloti bv. medicaginis/*M. laciniata*	STM 2832	C	B	O	R
	S. meliloti bv. medicaginis/*M. laciniata*	STM 2833	C	B	O	U
	S. meliloti bv. medicaginis/*M. laciniata*	STM 2834	C	A	E	X
	S. meliloti bv. medicaginis/*M. laciniata*	STM 2835	E	B	C	V
	S. meliloti bv. medicaginis/*M. laciniata*	STM 2836	C	B	O	W
	S. meliloti bv. medicaginis/*M. laciniata*	STM 2837	C	B	C	Z
	S. meliloti bv. medicaginis/*M. laciniata*	STM 2838	C	B	O	X
	S. meliloti bv. medicaginis/*M. laciniata*	STM 2839	C	B	E	Y
	S. meliloti bv. medicaginis/*M. laciniata*	STM 2840	C	B	O	X
	S. meliloti bv. medicaginis/*M. laciniata*	STM 2841	C	B	E	Y
Hadjeb	*S. meliloti* bv. meliloti/*M. truncatula*	STM 2842	C	B	F	K
	S. meliloti bv. meliloti/*M. truncatula*	STM 2843	A	D	E	D

Continued on following page

TABLE 3—*Continued*

Location	Taxon/host plant	Isolate	Allele at locus[a]			
			IGS_{RKP}	IGS_{EXO}	IGS_{GAB}	IGS_{NOD}
	S. meliloti bv. meliloti/*M. truncatula*	STM 2844	C	B	H	K
	S. meliloti bv. meliloti/*M. truncatula*	STM 2845	C	B	P	K
	S. meliloti bv. meliloti/*M. truncatula*	STM 2846	C	B	E	K
	S. meliloti bv. meliloti/*M. truncatula*	STM 2847	A	D	C	K
	S. meliloti bv. meliloti/*M. truncatula*	STM 2848	E	D	C	K
	S. meliloti bv. meliloti/*M. truncatula*	STM 2849	A	D	C	K
	S. meliloti bv. meliloti/*M. truncatula*	STM 2850	C	B	E	K
	S. meliloti bv. meliloti/*M. truncatula*	STM 2851	C	B	E	K
	S. meliloti bv. meliloti/*M. truncatula*	STM 2852	C	B	Q	K
	S. meliloti bv. meliloti/*M. truncatula*	STM 2853	A	D	C	K
	S. meliloti bv. meliloti/*M. truncatula*	STM 2854	C	B	O	A
	S. meliloti bv. meliloti/*M. truncatula*	STM 2855	C	B	C	K
	S. meliloti bv. meliloti/*M. truncatula*	STM 2856	C	B	O	A
	S. meliloti bv. meliloti/*M. truncatula*	STM 2857	C	B	O	K
	S. meliloti bv. meliloti/*M. truncatula*	STM 2858	C	B	R	K
	S. meliloti bv. meliloti/*M. truncatula*	STM 2859	C	B	E	A
	S. meliloti bv. meliloti/*M. truncatula*	STM 2860	C	B	C	K
	S. meliloti bv. meliloti/*M. truncatula*	STM 2861	C	B	H	L
	S. meliloti bv. meliloti/*M. truncatula*	STM 2862	C	B	E	B
	S. meliloti bv. meliloti/*M. truncatula*	STM 2863	A	B	C	K
	S. meliloti bv. meliloti/*M. truncatula*	STM 2864	C	B	O	B
	S. meliloti bv. meliloti/*M. truncatula*	STM 2865	C	B	O	K
	S. meliloti bv. meliloti/*M. truncatula*	STM 2866	G	B	E	K
	S. meliloti bv. meliloti/*M. truncatula*	STM 2867	A	D	E	D
	S. meliloti bv. meliloti/*M. truncatula*	STM 2868	C	B	E	A
	S. meliloti bv. meliloti/*M. truncatula*	STM 2869	C	B	E	A
	S. meliloti bv. meliloti/*M. truncatula*	STM 2870	C	B	H	K
	S. meliloti bv. meliloti/*M. truncatula*	STM 2871	C	B	O	A
	S. meliloti bv. meliloti/*M. truncatula*	STM 2872	C	B	E	A
	S. meliloti bv. meliloti/*M. truncatula*	STM 2873	A	B	C	K
	S. meliloti bv. meliloti/*M. truncatula*	STM 2874	A	D	E	E
	S. meliloti bv. meliloti/*M. truncatula*	STM 2875	C	B	O	K
	S. meliloti bv. meliloti/*M. truncatula*	STM 2876	C	B	C	A
	S. meliloti bv. meliloti/*M. truncatula*	STM 2877	H	D	C	K
	S. meliloti bv. meliloti/*M. truncatula*	STM 2878	B	B	E	C
	S. meliloti bv. meliloti/*M. truncatula*	STM 2879	C	B	O	B
	S. meliloti bv. meliloti/*M. truncatula*	STM 2880	C	B	C	K
	S. meliloti bv. meliloti/*M. truncatula*	STM 2881	C	C	E	K
	S. meliloti bv. meliloti/*M. truncatula*	STM 2882	A	B	C	K
	S. meliloti bv. meliloti/*M. truncatula*	STM 2883	C	B	C	K
	S. meliloti bv. medicaginis/*M. laciniata*	STM 2884	C	B	O	W
	S. meliloti bv. medicaginis/*M. laciniata*	STM 2885	C	B	B	U
	S. meliloti bv. medicaginis/*M. laciniata*	STM 2886	C	B	O	U
	S. meliloti bv. medicaginis/*M. laciniata*	STM 2887	C	B	O	Y
	S. meliloti bv. medicaginis/*M. laciniata*	STM 2888	C	B	E	Z
	S. meliloti bv. medicaginis/*M. laciniata*	STM 2889	C	B	F	T
	S. meliloti bv. medicaginis/*M. laciniata*	STM 2890	C	B	E	Y
	S. meliloti bv. medicaginis/*M. laciniata*	STM 2891	C	B	E	X
	S. meliloti bv. medicaginis/*M. laciniata*	STM 2892	E	D	B	X
	S. meliloti bv. medicaginis/*M. laciniata*	STM 2893	E	D	B	X
	S. meliloti bv. medicaginis/*M. laciniata*	STM 2894	C	B	O	X
	S. meliloti bv. medicaginis/*M. laciniata*	STM 2895	C	B	O	X
	S. meliloti bv. medicaginis/*M. laciniata*	STM 2896	C	B	I	X
	S. meliloti bv. medicaginis/*M. laciniata*	STM 2897	E	B	B	V
	S. meliloti bv. medicaginis/*M. laciniata*	STM 2898	C	B	E	X
	S. meliloti bv. medicaginis/*M. laciniata*	STM 2899	E	B	B	X
	S. meliloti bv. medicaginis/*M. laciniata*	STM 2900	C	B	E	Y
	S. meliloti bv. medicaginis/*M. laciniata*	STM 2901	E	D	B	X
	S. meliloti bv. medicaginis/*M. laciniata*	STM 2902	C	B	O	Y
	S. meliloti bv. medicaginis/*M. laciniata*	STM 2903	C	B	O	W

[a] Previously described alleles are identified by the same notation as that used by Bailly et al. (1).

the IGS_{RKP} data set. The low level of polymorphism within the last data set explained the choice of the simple JC69 model of evolution, since model choice is conditioned by the polymorphism observed at each locus. Statistics summarizing the polymorphism pattern for each sequence data set are presented in Table 4.

(ii) Population structure. We did not detect a significant overall differentiation between Enfidha and Hadjeb popula-

TABLE 4. Genetic diversity of *S. medicae*, *S. meliloti* bv. meliloti, and *S. meliloti* bv. medicaginis populations isolated from the three host plant groups used in this study

Sample and genomic region	No. of:			Genetic diversity[a]			
	Sequences	Aligned sites[b]	Segregating sites	*R* index		*H* index	
				Obs	N_{min}	Obs	N_{min}
S. medicae							
IGS_{RKP}	20	321 (0)	0	1	1.00 (1,1)	0.000	0.000 (0.000,0.000)
IGS_{NOD}	20	335 (0)	3	3	2.87 (2,3)	0.595	0.565 (0.426,0.668)
IGS_{GAB}	20	334 (3)	8	6	4.86 (3,6)	0.779	0.739 (0.647,0.816)
IGS_{EXO}	20	190 (0)	6	7	5.85 (4,7)	0.839	0.796 (0.711,0.858)
S. meliloti bv. meliloti							
IGS_{RKP}	66	302 (1)	6	7	3.65 (2,5)	0.480	0.475 (0.279,0.647)
IGS_{NOD}	66	349 (2)	7	8	4.31 (3,6)	0.544	0.549 (0.353,0.711)
IGS_{GAB}	66	274 (6)	15	13	6.50 (4,9)	0.799	0.789 (0.700,0.868)
IGS_{EXO}	66	241 (1)	11	3	2.46 (2,3)	0.369	0.373 (0.189,0.542)
S. meliloti bv. medicaginis							
IGS_{RKP}	40	302 (0)	4	3	2.37 (2,3)	0.337	0.323 (0.100,0.511)
IGS_{NOD}	40	415 (4)	24	10	6.94 (5,9)	0.823	0.803 (0.689,0.884)
IGS_{GAB}	40	274 (5)	11	6	4.86 (4,6)	0.749	0.732 (0.626,0.811)
IGS_{EXO}	40	241 (1)	11	3	2.26 (1,3)	0.229	0.219 (0.000,0.416)

[a] Obs, observed value. For N_{min}, the limits of the 95% confidence interval are given in parentheses. N_{min} values are set to 20.
[b] Number of insertion-deletion blocs within intraspecific alignments are provided in parentheses.

tions within each *S. meliloti* biovar ($P > 0.05$). Thus, Tunisian genotypes of each biovar isolated from the two soil samples were pooled in future analyses.

We first investigated the patterns of differentiation on the interspecific scale. No allele at any locus was shared between *S. medicae* and either *S. meliloti* biovar. We therefore observed high F_{ST} values between species, and all differentiation tests between these groups rejected the null hypothesis of no differentiation ($P < 0.001$). Moreover, large genetic divergences were estimated between the two species for the markers we used (Fig. 3A).

For IGS_{NOD}, while the genetic divergence between *S. meliloti* bv. medicaginis and *S. medicae* was almost the same as the divergence between the two *S. meliloti* biovars, it was 8.79 times higher than the divergence between *S. meliloti* bv. meliloti and *S. medicae*. Thus, for IGS_{NOD}, the divergence be-

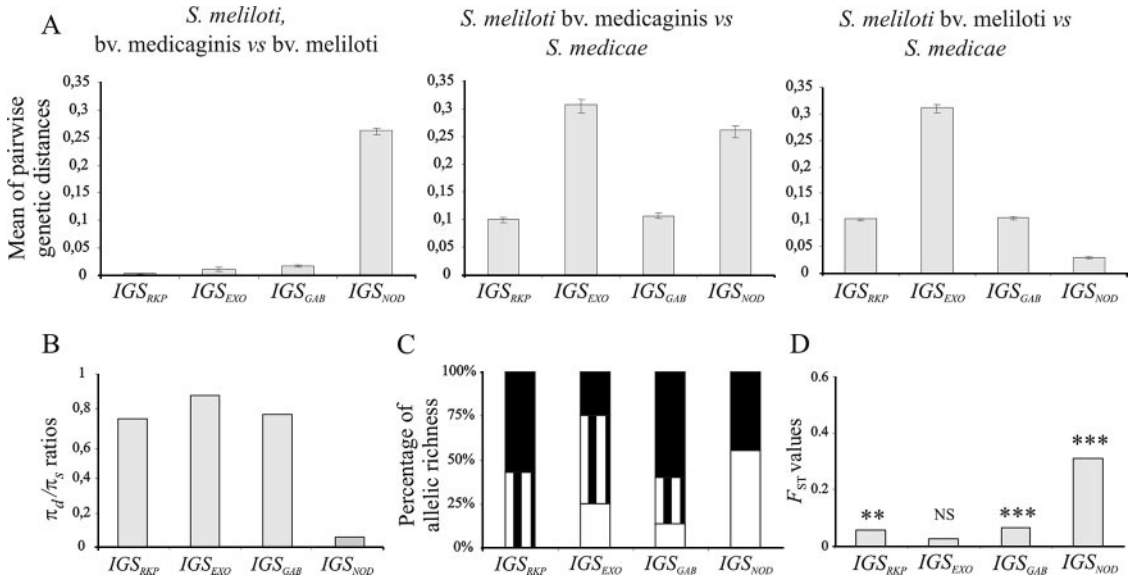

FIG. 3. Distributions of several measures illustrating the dissimilarity between population pairs belonging to either *S. meliloti* bv. meliloti, *S. meliloti* bv. medicaginis, or *S. medicae*. Data are provided independently for each of the four loci used for the MLSA. (A) Averages of the genetic distances among individuals belonging to two different bacterial groups (i.e., π_s), with 95% confidence intervals. (B) Ratios of the average genetic distances among individuals within and between *S. meliloti* biovars (i.e., π_d/π_s). (C) Percentages of allelic richness specific to *S. meliloti* bv. meliloti and *S. meliloti* bv. medicaginis, indicated by black and white areas, respectively. Striped areas illustrate the percentages of allelic richness shared by pairs of taxa. (D) Distribution of pairwise F_{ST} values between *S. meliloti* biovars. NS, **, and ***, differentiation test *P* values above a 5% threshold, below a 1% threshold, and below a 0.1% threshold, respectively.

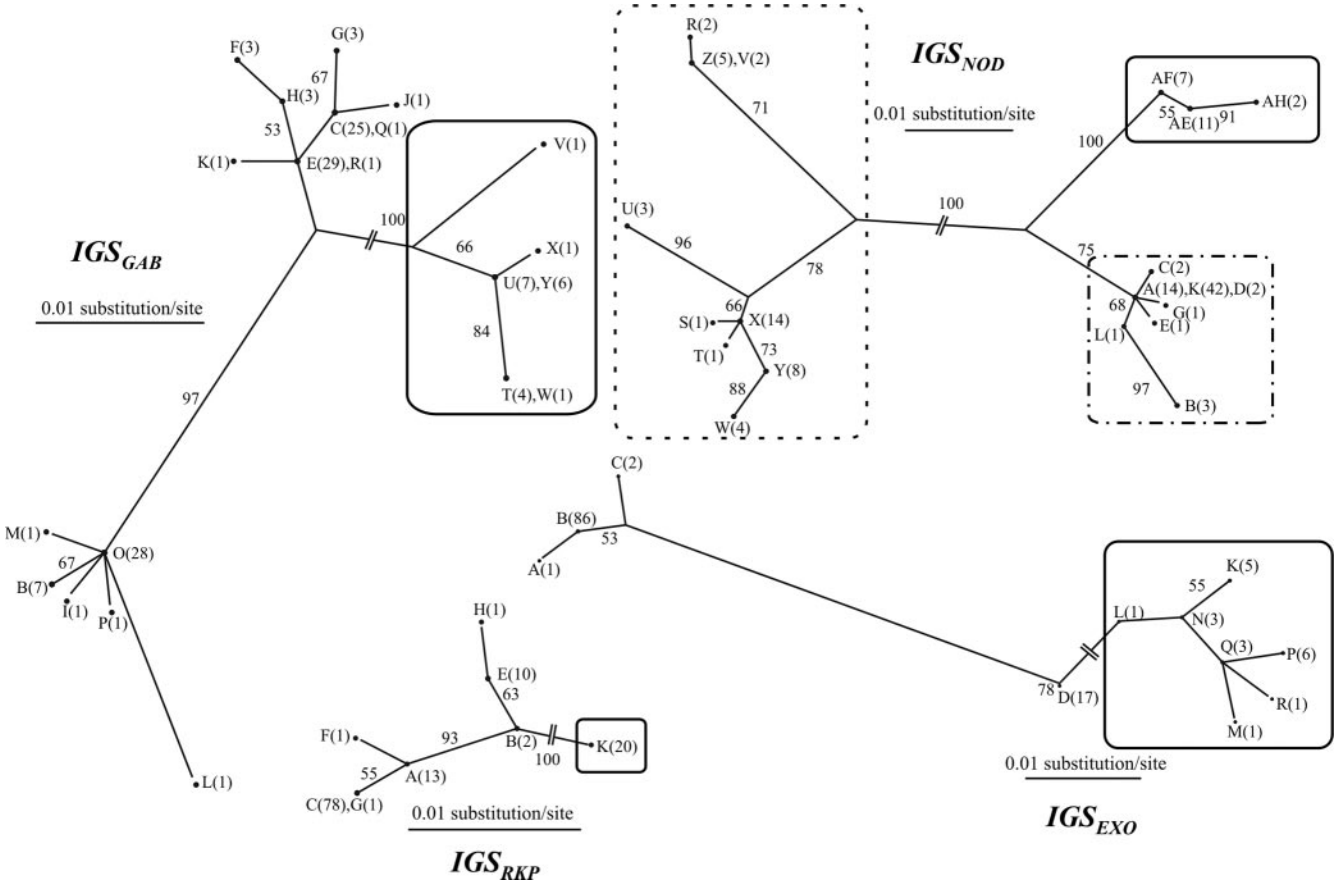

FIG. 4. Maximum likelihood trees obtained for each of the four markers used in the MLSA scheme. The number of occurrences of each allele of the data set is indicated in parentheses. Alleles belonging to either *S. medicae*, *S. meliloti* bv. medicaginis, or *S. meliloti* bv. meliloti isolates are framed by solid lines, a dotted line, and a dashed-dotted line, respectively, if they group in a clade. A double-dash symbol on a branch indicates that the branch is not drawn to scale (see Fig. 3 for more information about genetic divergence between bacterial groups). *S. medicae* alleles cluster in a clade on all phylogenies. This is consistent with the hypothesis of sexual isolation between *S. meliloti* and *S. medicae*. Conversely, isolates belonging to either *S. meliloti* bv. meliloti or *S. meliloti* bv. medicaginis do not cluster in monophyletic groups in all phylogenies except that for IGS_{NOD}. The incongruence among the four topologies suggests that recombination occurred among *S. meliloti* biovars.

tween biovars within a species was much larger than the divergence between one of the two biovars and *S. medicae*. This intriguing pattern was consistent with the particular phylogenetic relationship among the three groups inferred from the IGS_{NOD} maker. IGS_{NOD} alleles of both *S. medicae* and *S. meliloti* bv. meliloti indeed grouped in a monophyletic cluster, whereas *S. meliloti* bv. medicaginis alleles formed a sister clade. This is also consistent with the results we obtained based on the *nodABC-nodEG* phylogeny (Fig. 2).

Conversely, the divergence between the two *S. meliloti* biovars was 5 to 16 times lower than the divergence between *S. meliloti* and *S. medicae* at IGS_{EXO} (16.06 times), IGS_{RKP} (16.56 times), and IGS_{GAB} (5.22 times) (Fig. 4). Therefore, for these three loci, the divergence between biovars within a species was much lower than the divergence between any biovar and the other species.

On the intraspecific scale, *S. meliloti* bv. meliloti and *S. meliloti* bv. medicaginis indeed shared several alleles for all loci but IGS_{NOD} (Fig. 3C and 4). As a consequence, the F_{ST} value observed for IGS_{NOD} was much higher than those computed for the other loci, even though the two biovars were significantly differentiated at all loci but IGS_{EXO} (Fig. 3D). Furthermore, the mean molecular divergence between the two *S. meliloti* biovars was low for all loci but IGS_{NOD} (Fig. 3A). In the same way, the π_d/π_s ratio, i.e., the mean pairwise divergence between isolates that belong to two different biovars of *S. meliloti* over the mean pairwise divergence within these groups, was much lower for IGS_{NOD} than for the other loci (Fig. 3B). At this point, it is noteworthy that IGS_{NOD} is the only marker supporting a topology congruent with the delineation of the two *S. meliloti* biovars and, therefore, the evolution toward two specificity groups within this species. For IGS_{RKP}, IGS_{EXO}, and IGS_{GAB}, Shimodaira-Hasegawa tests comparing the likelihoods of data sets based on constrained (i.e., enforcing each *S. meliloti* biovar from a monophyletic group) and unconstrained topologies revealed significant differences ($P < 0.05$).

(iii) Recombination evidence. Since we detected obvious differences in the population structures revealed by the four MLSA loci on the intraspecific scale, we looked for traces of recombination events within our data set. Since significant differentiation had been detected among *S. meliloti* biovars, we used a phylogenetic approach rather than population genetic

tools such as linkage disequilibrium to study the influence of sex on *S. meliloti* diversification. The last measures indeed depend on both the sexual behavior and the genetic structure of the population under study. Furthermore, all of the Shimodaira-Hasegawa tests which aimed at comparing the likelihoods of the data based on the four phylogenies obtained from MLSA markers rejected the null hypothesis, suggesting that IGS_{GAB}, IGS_{EXO}, and IGS_{RKP} also evolved independently ($P < 0.05$). Finally, based on the scan of the concatenated sequence alignment of the four loci with MaxChi, Chimaera, and Geneconv, we consistently inferred recombination breakpoints ($P < 0.05$) located between each locus by using sequence sets, i.e., pairs or triplets, which included genotypes belonging to either one or two *S. meliloti* biovars.

DISCUSSION

Origin of the *nod* genes of *Sinorhizobium meliloti* and *S. medicae*. We obtained a supertree including all known specificity groups of *Medicago* symbionts, based on the partial sequences of the *nodABC-nodEG* gene clusters. This phylogeny included *S. medicae* and three ecotypes of *S. meliloti*, which are the regular symbionts of most *Medicago* species. It also included *Rhizobium mongolense*, which is described as the specific symbiont of *Medicago ruthenica* (53), despite a few strains having been isolated from nodules of plant genera characterized by their low symbiotic specificity, such as *Hedysarum* (4, 43, 47, 60). This phylogeny was rooted using an outgroup composed of a "core collection" of the diverse *nod* genes sampled from strains belonging to the *Rhizobiaceae* family (Fig. 2). The topology we obtained is congruent with other *nod* trees inferred from single gene sequences and is especially supportive of two main hypotheses, i.e., (i) the monophyly of *R. leguminosarum* bv. viciae, *R. leguminosarum* bv. trifolii, *R. galegae*, and *Medicago* symbionts based on *nodC* and *nodA* sequences (23, 50); and (ii) the monophyly of the cluster composed of *R. mongolense* isolates and *Sinorhizobium* strains associated with *Medicago* based on *nodB* sequences (43).

These clustering patterns are remarkably related to those inferred from the phylogeny of legumes (61). *Galega* species, which interact with *R. galegae* (35), form a sister group to the vicoid clade. This clade includes the following three main lineages: (i) the Vicieae tribe (i.e., *Vicia*, *Lens*, *Lathyrus*, and *Pisum* spp.), whose members interact with *R. leguminosarum* bv. viciae (32); (ii) the genus *Trifolium*, which is associated with *R. leguminosarum* bv. trifolii (46); and (iii) other members of the Trifolieae tribe (i.e., *Ononis*, *Trigonella*, *Melilotus*, and *Medicago* spp.), which mostly interact with *S. meliloti* and *S. medicae*, with the only obvious exceptions to this symbiotic specificity being the genus *Ononis*, which interacts with a much wider range of strains (46), and the species *Medicago ruthenica*, which interacts with *R. mongolense*. Nevertheless, it is noteworthy that *M. ruthenica* appears to be the first divergent plant species in the strict consensus most-parsimonious molecular phylogeny of the genus *Medicago* (13). In accordance with phylogenetic evidence, *Galega* species and vicoid species are known to share the peculiar ability to interact with rhizobia that produce Nod factors (i.e., the products of the *nod* gene pathway) with alpha-beta-unsaturated fatty acids, such as *R. galegae*, *R. leguminosarum* bv. viciae, *R. leguminosarum* bv. trifolii, or *S. meliloti* (12). These data suggest that coevolution played a fundamental role in the diversification of both the symbiotic specificity of vicoid species and the *nod* genes of their specific symbionts, even though a strict cocladogenesis was not observed.

Conversely, phylogenies based on chromosomal genes do not support the monophyly of the bacterial lineages which interact with *Medicago* spp., with the topologies being consistent with the monophyly of the *Sinorhizobium* and *Rhizobium* genera (e.g., see reference 52). The incongruence between symbiotic and housekeeping gene phylogenies suggests that HGT of *nod* genes occurred between these two genera. In this context, the analyses performed up to now suggest that *Sinorhizobium* species acquired the ability to nodulate *Medicago* species through HGT of *nod* genes from the ancestor of a *Rhizobium* species, most probably *R. mongolense*. Such HGT across different taxonomic groups of bacteria is usually invoked to explain the spread of *nod* genes, both between alpha- and betaproteobacteria (7) and within alphaproteobacteria (60). This hypothesis was validated, for instance, with the report of transfer of nodulation and N_2 fixation genes by acquisition of a symbiotic island that integrated into a Phe-tRNA that occurred in natura between *Mesorhizobium* strains (48).

Speciation and specialization of *S. meliloti* and *S. medicae*. During the evolution of *Medicago* symbionts, HGT of *nod* genes might have occurred either towards the common ancestor of the two *Sinorhizobium* species or towards one *Sinorhizobium* species followed by a spread to the other. All the topologies we obtained based on molecular markers located in the *nod* gene region showed that the sequences of *S. meliloti* bv. meliloti and *S. medicae* cluster in a common clade, while the whole *S. meliloti* species forms a paraphyletic group. Moreover, *S. meliloti* bv. meliloti also forms a paraphyletic group on the *nodABC-nodEG* supertree due to the position of the *S. medicae* clade. This branching pattern is explained by the incomplete lineage sorting of *S. meliloti* bv. meliloti and *S. medicae* at the *nodA* and *nodB* genes and is in agreement with topologies previously obtained based on *nodD1* or *nodB* sequences (3, 42).

The close relationship between the *nod* genes of *S. meliloti* bv. meliloti and *S. medicae* is consistent with the fact that these two groups are characterized by overlapping host ranges (3). Conversely, the differentiation and divergence we observed for most MLSA loci between *S. medicae* and *S. meliloti* illustrate a substantial process of sexual isolation, consistent with previous results illustrating the differentiation between *S. medicae* and *S. meliloti* bv. meliloti (1, 14, 31, 54). Moreover, phylogenies based on a sample of housekeeping genes suggest that *S. meliloti* and *S. medicae* are not sister species (28). Most of the topologies obtained indeed suggest that *S. medicae* might be the first emerging taxon within a clade including *S. medicae* and *S. meliloti* but also another species, *Sinorhizobium arboris* (28), suggesting a rather ancient speciation event leading to the first two species.

Two different scenarios might explain the contrast between the polymorphism patterns observed for *nod* genes and other markers. First, horizontal transfer of *nod* genes might have occurred between *S. meliloti* and *S. medicae* after their speciation. According to this hypothesis, the close genetic relationship between *nod* genes of *S. medicae* and *S. meliloti* bv. me-

liloti should be due to their recent divergence. An alternative explanation which has previously been proposed is that *S. medicae* emerged from an ancestral population of *S. meliloti* bv. meliloti (6, 42). According to this scenario, the monophyly of the *nod* sequences of *S. medicae* and *S. meliloti* bv. meliloti would be due to historical factors, and after speciation, selective pressures imposed on the *nod* gene cluster by the host plants shared by the two groups could have prevented their divergence. While the divergence between *S. medicae* and *S. meliloti* bv. meliloti was significantly less for *nod* genes than for a sample of housekeeping genes, we did not detect a significant difference in the K_a/K_s ratios for these two groups of genes, suggesting that selection influenced sequence divergence in a similar way for the two groups of genes. Whereas the two groups of genes evolved under purifying selective pressures, as revealed by K_a/K_s ratios of <1, the similarity of K_a/K_s ratios of *nod* genes and housekeeping genes suggests that purifying selection driven by host plants could not explain the low level of divergence between *S. medicae* and *S. meliloti* bv. meliloti in the *nod* gene cluster.

Altogether, these results suggest that HGT of *nod* genes between *S. medicae* and *S. meliloti* bv. meliloti is the most likely explanation for the pattern we observed. Moreover, the discrepancy between the topologies supported by *nodD1-nodAB* and *nodEG* raises questions about the number of transfers that occurred between the two groups. Further insights about these issues will likely be obtained from comparison of the genome sequences of *S. meliloti* bv. meliloti strain 1021 and *S. medicae* strain WSM419 when the genome of the latter strain becomes available.

Recombination within *S. meliloti* and ecological specialization. Previous studies have demonstrated that the genetic structure of *S. meliloti* bv. meliloti is strongly influenced by homologous recombination (1, 31, 49, 54). Nevertheless, this gives us no direct information about the impact of recombination on the evolution of bacterial populations in the context of ecological specialization. Based on current knowledge of the bacterial speciation process, the impact of disruptive selection on the pattern of bacterial diversification remains an open question (15). In particular, some models predict that global selective sweeps affecting all conspecific ecotypes are required to limit the genetic divergence of the niche-specific lineages (9). Here we analyzed the genetic structure of *S. meliloti* biovars adapted to different host plants in order to know whether homologous recombination could prevent conspecific bacterial ecotypes from undergoing speciation. With this aim, we summarize the evidence of disruptive selection within the *nod* gene cluster of *S. meliloti* bv. meliloti and *S. meliloti* bv. medicaginis and then focus on the recombination mechanism that would affect this set of genes. Next, the impact of recombination and disruptive selection on the diversification of other genomic areas is considered by comparing the observed patterns of polymorphism with the expected patterns provided by a theoretical model which was previously published. Finally, the ecological conditions required to allow recombination among biovars are discussed.

S. meliloti bv. meliloti and *S. meliloti* bv. medicaginis are completely differentiated at IGS_{NOD} (i.e., they have no shared allele). Moreover, IGS_{NOD} was the only marker supporting the split between the two *S. meliloti* biovars, based on Shimodaira-Hasegawa tests. According to the model proposed by Majewski et al. (26), the π_s/π_d ratio, i.e., the sequence divergence within ecotypes over the sequence divergence between ecotypes, for IGS_{NOD} (i.e., 5.85×10^{-2}) would be obtained in cases where recombination is not strong enough to prevent the sorting of ecotypes. We therefore believe that a disruptive selection pressure due to the host plant has influenced the variability of some *S. meliloti nod* genes and might have further modified the polymorphism pattern of the entire *nod* gene region by hitchhiking. This hypothesis is consistent with a previous study which revealed that the efficiency of the interaction of *S. meliloti* bv. medicaginis strains with *M. laciniata* required a specific *nodC* allele (2). Furthermore, the high level of divergence between *S. meliloti* ecotypes at IGS_{NOD} raises questions about the mechanism of recombination that affects the *nod* gene cluster. According to previous studies showing that the efficiency of homologous recombination depends on sequence divergence in a log-linear way (27, 58), the recombination rate among *nod* gene clusters of the different biovars of *S. meliloti* would actually be quite low.

Nevertheless, Shimodaira-Hasegawa tests show that recombination uncoupled the evolution of IGS_{NOD} and IGS_{GAB}, both on pSymA. This agrees with results from a previous study in which bacterial sampling was restricted to a single soil sample and a single biovar, where no significant linkage disequilibrium was found between these two molecular markers (1). Conversely, Sun et al. (50) found some significant linkage disequilibrium among the markers they studied on pSymA (49). Yet *nod* gene polymorphism was not taken into account in their study. According to the evidence of recombination between IGS_{NOD} and IGS_{GAB}, the lack of significant evidence of recombination within the *nodABC-nodEG* gene cluster, and Sun et al.'s results (i.e., linkage disequilibrium among pSymA loci that are not located within the *nod* gene cluster), it can be hypothesized that the whole *nod* gene region is frequently transferred among *S. meliloti* isolates and behaves as a hot spot of homologous recombination within the megaplasmid pSymA of *S. meliloti*. In agreement with this hypothesis, a study based on comparative genomic hybridization of *S. meliloti* strains revealed that pSymA is a hot spot of genomic polymorphism within the *S. meliloti* genome and that a gain or loss of genes tends to be clustered spatially in this replication unit, especially near *nod* genes (18). Alternatively, it could be that the linkage disequilibria found by Sun et al. (49) were essentially due to population structure, as the strains they studied originated from all over the world.

Furthermore, the significant incongruence among IGS_{RKP}, IGS_{EXO}, IGS_{GAB}, and IGS_{NOD} phylogenies confirmed that homologous recombination has influenced the evolution of *S. meliloti* for the different markers we used. Recombination analyses allowed us to infer homologous recombination events between *S. meliloti* bv. meliloti and *S. meliloti* bv. medicaginis, showing that they are not sexually isolated. Lastly, we observed for IGS_{RKP}, IGS_{EXO}, and IGS_{GAB} that π_s and π_d are almost the same, with the ratio of the two ranging from 7.32×10^{-1} to 8.64×10^{-1}. According to Majewski and Cohan's theoretical model on the effect of ecological specialization on bacterial diversification (26), such a pattern is expected when the rate of recombination on the whole-species scale is the main factor that limits the divergence between ecotypes. In spite of disrup-

tive selective pressure, sex makes the two groups poorly distinguishable for molecular markers not involved in ecological specialization. This suggests that divergence at *nod* loci between biovars neither prevents the transfer of this genomic area nor leads to an overall decrease of homologous recombination between *S. meliloti* genomes. It also suggests that no strong coevolution between genes involved in symbiotic specificity, in particular *nod* genes, and the core genome constrains gene exchanges among ecotypes. Within *S. meliloti*, the efficiency of *nod* gene variants would not be affected by the genetic background in which they are located.

Finally, to allow homologous recombination to occur, both *S. meliloti* biovars have to mate under suitable ecological conditions. Pretorius-Guth et al. (34) suggested that plant nodules create the most favorable natural conditions for recombination among *S. meliloti* isolates. Under this hypothesis, the maintenance of a common genetic background might thus involve coinfection of both *S. meliloti* biovars within mixed nodules, eluding the barrier to gene flow due to host plant specificity (17). Such events could indeed occur in natura since the coinoculation of nodulation mutants (Nod^-) and infection mutants deficient in exopolysaccharide production (Inf^- EPS^-) during ex situ experiments results in mixed infected nodules (22). For such coinfection to occur, it is necessary that the host plant species of both biovars of *S. meliloti* cooccur locally, which is true for *M. truncatula* and *M. laciniata* (50). Alternatively, frequent disturbance events that would homogenize the rhizosphere could also facilitate such coinoculation events, and thus gene flow among biovars.

ACKNOWLEDGMENTS

We thank M. E. Aouani for providing us with Tunisian soil samples. We also thank the Australian *Medicago* Genetic Resource Centre for providing us with the bulk of Tunisian genotypes of *Medicago laciniata*. Lucette Maure and Gaëlle Leonard provided technical help for strain isolation. We acknowledge two anonymous reviewers for their constructive and helpful comments.

This research was funded by the Bureau des Ressources Génétiques, Programme Recherches Méthodologiques pour l'Amélioration des Processus de Gestion et de Conservation des Ressources Génétiques Animales, Végétales et Microbiennes, to I.O. G. B. X. Bailly was supported by a Ph.D. fellowship from the French Ministry of Education and Research and by a Jean and Marie-Louise Dufrenoy fellowship from the French Academy of Agriculture.

This is publication ISEM 2007.039 of the Institut des Sciences de l'Evolution, Montpellier, France.

REFERENCES

1. **Bailly, X., I. Olivieri, S. De Mita, J. C. Cleyet-Marel, and G. Bena.** 2006. Recombination and selection shape the molecular diversity pattern of nitrogen-fixing *Sinorhizobium* sp. associated to *Medicago*. Mol. Ecol. **15:**2719–2734.
2. **Barran, L. R., E. S. Bromfield, and D. C. Brown.** 2002. Identification and cloning of the bacterial nodulation specificity gene in the *Sinorhizobium meliloti-Medicago laciniata* symbiosis. Can. J. Microbiol. **48:**765–771.
3. **Bena, G., A. Lyet, T. Huguet, and I. Olivieri.** 2005. *Medicago-Sinorhizobium* symbiotic specificity evolution and the geographic expansion of *Medicago*. J. Evol. Biol. **18:**1547–1558.
4. **Benhizia, Y., H. Benhizia, A. Benguedouar, R. Muresu, A. Giacomini, and A. Squartini.** 2004. Gamma proteobacteria can nodulate legumes of the genus *Hedysarum*. Syst. Appl. Microbiol. **27:**462–468.
5. **Bininda-Emonds, O. R.** 2004. The evolution of supertrees. Trends Ecol. Evol. **19:**315–322.
6. **Biondi, E. G., E. Pilli, E. Giuntini, M. L. Roumiantseva, E. E. Andronov, O. P. Onichtchouk, O. N. Kurchak, B. V. Simarov, N. I. Dzyubenko, A. Mengoni, and M. Bazzicalupo.** 2003. Genetic relationship of *Sinorhizobium meliloti* and *Sinorhizobium medicae* strains isolated from Caucasian region. FEMS Microbiol. Lett. **220:**207–213.
7. **Chen, W. M., L. Moulin, C. Bontemps, P. Vandamme, G. Bena, and C. Boivin-Masson.** 2003. Legume symbiotic nitrogen fixation by beta-proteobacteria is widespread in nature. J. Bacteriol. **185:**7266–7272.
8. **Cohan, F. M.** 2001. Bacterial species and speciation. Syst. Biol. **50:**513–524.
9. **Cohan, F. M.** 2002. What are bacterial species? Annu. Rev. Microbiol. **56:**457–487.
10. **Criscuolo, A., V. Berry, E. J. Douzery, and O. Gascuel.** 2006. SDM: a fast distance-based approach for (super)tree building in phylogenomics. Syst. Biol. **55:**740–755.
11. **Daubin, V., M. Gouy, and G. Perriere.** 2002. A phylogenomic approach to bacterial phylogeny: evidence of a core of genes sharing a common history. Genome Res. **12:**1080–1090.
12. **Debelle, F., L. Moulin, B. Mangin, J. Denarie, and C. Boivin.** 2001. Nod genes and Nod signals and the evolution of the Rhizobium legume symbiosis. Acta Biochim. Pol. **48:**359–365.
13. **Downie, S. R., D. Katz-Downie, E. J. Rogers, H. L. Zujewski, and E. Small.** 1998. Multiple independent losses of the plastid *rpoC1* intron in *Medicago* (Fabaceae) as inferred from phylogenetic analyses of nuclear ribosomal DNA internal transcribed spacer sequences. Can. J. Bot. **76:**791–803.
14. **Eardly, B. D., L. A. Materon, N. H. Smith, D. A. Johnson, M. D. Rumbaugh, and R. K. Selander.** 1990. Genetic structure of natural populations of the nitrogen-fixing bacterium *Rhizobium meliloti*. Appl. Environ. Microbiol. **56:**187–194.
15. **Fraser, C., W. P. Hanage, and B. G. Spratt.** 2007. Recombination and the nature of bacterial speciation. Science **315:**476–480.
16. **Galibert, F., et al.** 2001. The composite genome of the legume symbiont *Sinorhizobium meliloti*. Science **293:**668–672.
17. **Gavrilets, S.** 2003. Perspective: models of speciation: what have we learned in 40 years? Evolution **57:**2197–2215.
18. **Giuntini, E., A. Mengoni, C. De Filippo, D. Cavalieri, N. Aubin-Horth, C. R. Landry, A. Becker, and M. Bazzicalupo.** 2005. Large-scale genetic variation of the symbiosis-required megaplasmid pSymA revealed by comparative genomic analysis of *Sinorhizobium meliloti* natural strains. BMC Genomics **6:**158.
19. **Hall, T. A.** 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. **41:**95–98.
20. **Huelsenbeck, J. P., and F. Ronquist.** 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics **17:**754–755.
21. **Jain, R., M. C. Rivera, J. E. Moore, and J. A. Lake.** 2003. Horizontal gene transfer accelerates genome innovation and evolution. Mol. Biol. Evol. **20:**1598–1602.
22. **Kapp, D., K. Niehaus, J. Quandt, P. Muller, and A. Puhler.** 1990. Cooperative action of *Rhizobium meliloti* nodulation and infection mutants during the process of forming mixed infected alfalfa nodules. Plant Cell **2:**139–151.
23. **Laguerre, G., S. M. Nour, V. Macheret, J. Sanjuan, P. Drouin, and N. Amarger.** 2001. Classification of rhizobia based on *nodC* and *nifH* gene analysis reveals a close phylogenetic relationship among *Phaseolus vulgaris* symbionts. Microbiology **147:**981–993.
24. **Lawrence, J. G.** 2002. Gene transfer in bacteria: speciation without species? Theor. Popul. Biol. **61:**449–460.
25. **Maiden, M. C., J. A. Bygraves, E. Feil, G. Morelli, J. E. Russell, R. Urwin, Q. Zhang, J. Zhou, K. Zurth, D. A. Caugant, I. M. Feavers, M. Achtman, and B. G. Spratt.** 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc. Natl. Acad. Sci. USA **95:**3140–3145.
26. **Majewski, J., and F. M. Cohan.** 1999. Adapt globally, act locally: the effect of selective sweeps on bacterial sequence diversity. Genetics **152:**1459–1474.
27. **Majewski, J., and F. M. Cohan.** 1999. DNA sequence similarity requirements for interspecific recombination in *Bacillus*. Genetics **153:**1525–1533.
28. **Martens, M., M. Delaere, R. Coopman, P. De Vos, M. Gillis, and A. Willems.** 2007. Multilocus sequence analysis of *Ensifer* and related taxa. Int. J. Syst. Evol. Microbiol. **57:**489–503.
29. **Martin, D. P., C. Williamson, and D. Posada.** 2005. RDP2: recombination detection and analysis from sequence alignments. Bioinformatics **21:**260–262.
30. **Martin, G., S. P. Otto, and T. Lenormand.** 2006. Selection for recombination in structured populations. Genetics **172:**593–609.
31. **Maynard Smith, J., N. H. Smith, M. O'Rourke, and B. G. Spratt.** 1993. How clonal are bacteria? Proc. Natl. Acad. Sci. USA **90:**4384–4388.
32. **Mutch, L. A., and J. P. Young.** 2004. Diversity and specificity of *Rhizobium leguminosarum* biovar viciae on wild and cultivated legumes. Mol. Ecol. **13:**2435–2444.
33. **Nei, M.** 1987. Molecular evolutionary genetics. Columbia University Press, New York, NY.
34. **Pretorius-Guth, I. M., A. Puhler, and R. Simon.** 1990. Conjugal transfer of megaplasmid 2 between *Rhizobium meliloti* strains in alfalfa nodules. Appl. Environ. Microbiol. **56:**2354–2359.
35. **Radeva, G., G. Jurgens, M. Niemi, G. Nick, L. Suominen, and K. Lindstrom.** 2001. Description of two biovars in the *Rhizobium galegae* species: biovar Orientalis and biovar Officinalis. Syst. Appl. Microbiol. **24:**192–205.

36. **Raymond, M., and F. Rousset.** 1995. An exact test for population differentiation. Evolution **49:**1280–1283.
37. **Raymond, M., and F. Rousset.** 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. J. Hered. **86:**248–249.
38. **Rome, S., M. P. Fernandez, B. Brunel, P. Normand, and J. C. Cleyet-Marel.** 1996. *Sinorhizobium medicae* sp. nov., isolated from annual *Medicago* spp. Int. J. Syst. Bacteriol. **46:**972–980.
39. **Rozas, J., J. C. Sanchez-DelBarrio, X. Messeguer, and R. Rozas.** 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics **19:**2496–2497.
40. **Sawyer, S.** 1989. Statistical tests for detecting gene conversion. Mol. Biol. Evol. **6:**526–538.
41. **Shimodaira, H., and M. Hasegawa.** 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. **16:**1114–1116.
42. **Silva, C., F. L. Kan, and E. Martinez-Romero.** 2007. Population genetic structure of *Sinorhizobium meliloti* and *S. medicae* isolated from nodules of *Medicago* spp. in Mexico. FEMS Microbiol. Ecol. **60:**477–489.
43. **Silva, C., P. Vinuesa, L. E. Eguiarte, V. Souza, and E. Martinez-Romero.** 2005. Evolutionary genetics and biogeographic structure of *Rhizobium gallicum* sensu lato, a widely distributed bacterial symbiont of diverse legumes. Mol. Ecol. **14:**4033–4050.
44. **Small, E., and M. Jomphe.** 1988. A synopsis of the genus *Medicago* (Leguminosae). Can. J. Bot. **67:**3260–3294.
45. **Smets, B. F., and T. Barkay.** 2005. Horizontal gene transfer: perspectives at a crossroads of scientific disciplines. Nat. Rev. Microbiol. **3:**675–678.
46. **Sprent, J.** 2001. Nodulation in legumes. Royal Botanic Gardens, Kew, United Kingdom.
47. **Squartini, A., P. Struffi, H. Doring, S. Selenska-Pobell, E. Tola, A. Giacomini, E. Vendramin, E. Velazquez, P. F. Mateos, E. Martinez-Molina, F. B. Dazzo, S. Casella, and M. P. Nuti.** 2002. *Rhizobium sullae* sp. nov. (formerly '*Rhizobium hedysari*'), the root-nodule microsymbiont of *Hedysarum coronarium* L. Int. J. Syst. Evol. Microbiol. **52:**1267–1276.
48. **Sullivan, J. T., and C. W. Ronson.** 1998. Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into a Phe-tRNA gene. Proc. Natl. Acad. Sci. USA **95:**5145–5149.
49. **Sun, S., H. Guo, and J. Xu.** 2006. Multiple gene genealogical analyses reveal both common and distinct population genetic patterns among replicons in the nitrogen-fixing bacterium *Sinorhizobium meliloti*. Microbiology **152:**3245–3259.
50. **Suominen, L., C. Roos, G. Lortet, L. Paulin, and K. Lindstrom.** 2001. Identification and structure of the *Rhizobium galegae* common nodulation genes: evidence for horizontal gene transfer. Mol. Biol. Evol. **18:**907–916.
51. **Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins.** 1997. The ClustalX Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **25:**4876–4882.
52. **Turner, S. L., and J. P. Young.** 2000. The glutamine synthetases of rhizobia: phylogenetics and evolutionary implications. Mol. Biol. Evol. **17:**309–319.
53. **Van Berkum, P., D. Beyene, G. Bao, T. A. Campbell, and B. D. Eardly.** 1998. *Rhizobium mongolense* sp. nov. is one of the three rhizobial genotypes identified which nodulate and form nitrogen fixing symbiosis with *Medicago ruthenica* [(L.) Ledebour]. Int. J. Syst. Bacteriol. **48:**13–22.
54. **Van Berkum, P., P. Elia, and B. D. Eardly.** 2006. Multilocus sequence typing as an approach for population analysis of *Medicago*-nodulating rhizobia. J. Bacteriol. **188:**5570–5577.
55. **Vetsigian, K., and N. Goldenfeld.** 2005. Global divergence of microbial genome sequences mediated by propagating fronts. Proc. Natl. Acad. Sci. USA **102:**7332–7337.
56. **Villegas, M. D. C., S. Rome, L. Maure, O. Domergue, L. Gardan, X. Bailly, J. C. Cleyet-Marel, and B. Brunel.** 2006. Nitrogen-fixing sinorhizobia with *Medicago laciniata* constitute a novel biovar (bv. medicaginis) of *S. meliloti*. Syst. Appl. Microbiol. **29:**526–538.
57. **Vincent, J. M.** 1970. A manual for the practical study of root-nodule bacteria, vol. 15. Blackwell Scientific Publications Ltd., Oxford, United Kingdom.
58. **Vulic, M., F. Dionisio, F. Taddei, and M. Radman.** 1997. Molecular keys to speciation: DNA polymorphism and the control of genetic exchange in enterobacteria. Proc. Natl. Acad. Sci. USA **94:**9763–9767.
59. **Weir, B. S., and C. C. Cockerham.** 1984. Estimating F-statistics for the analysis of population structure. Evolution **38:**1358–1370.
60. **Wernegreen, J. J., and M. A. Riley.** 1999. Comparison of the evolutionary dynamics of symbiotic and housekeeping loci: a case for the genetic coherence of rhizobial lineages. Mol. Biol. Evol. **16:**98–113.
61. **Wojciechowski, M. F., M. J. Sanderson, K. P. Steele, and A. Liston.** 2000. Molecular phylogeny of the "temperate herbaceous tribes" of papilionoid legumes: a supertree approach, p. 277–298. *In* P. S. Herendeen and A. Bruneau (ed.), Advances in legume systematics 9. Royal Botanic Gardens, Kew, United Kingdom.