Microbiology (2001), **147**, 981–993 Printed in Great Britain

Classification of rhizobia based on *nodC* and *nifH* gene analysis reveals a close phylogenetic relationship among *Phaseolus vulgaris* symbionts

Gisèle Laguerre,[1] Sarah M. Nour,[1]† Valérie Macheret,[1] Juan Sanjuan,[2] Pascal Drouin[1]‡ and Noëlle Amarger[1]

Author for correspondence: Gisèle Laguerre. Tel: +33 3 80693093. Fax: +33 3 80693224. e-mail: laguerre@dijon.inra.fr

[1] Laboratoire de Microbiologie des Sols, Centre de Microbiologie du Sol et de l'Environnement, INRA, 17 rue Sully, BP 86510, F-21065 Dijon Cedex, France

[2] Departamento de Microbiología del Suelo y Sistemas Simbióticos, Estación Experimental del Zaidín, CSIC, Profesor Albareda 1, E-18008 Granada, Spain

The *nodC* and *nifH* genes were characterized in a collection of 83 rhizobial strains which represented 23 recognized species distributed in the genera *Rhizobium*, *Sinorhizobium*, *Mesorhizobium* and *Bradyrhizobium*, as well as unclassified rhizobia from various host legumes. Conserved primers were designed from available nucleotide sequences and were able to amplify *nodC* and *nifH* fragments of about 930 bp and 780 bp, respectively, from most of the strains investigated. RFLP analysis of the PCR products resulted in a classification of these rhizobia which was in general well-correlated with their known host range and independent of their taxonomic status. The *nodC* and *nifH* fragments were sequenced for representative strains belonging to different genera and species, most of which originated from *Phaseolus vulgaris* nodules. Phylogenetic trees were constructed and revealed close relationships among symbiotic genes of the *Phaseolus* symbionts, irrespective of their 16S-rDNA-based classification. The *nodC* and *nifH* phylogenies were generally similar, but cases of incongruence were detected, suggesting that genetic rearrangements have occurred in the course of evolution. The results support the view that lateral genetic transfer across rhizobial species and, in some instances, across *Rhizobium* and *Sinorhizobium* genera plays a role in diversification and in structuring the natural populations of rhizobia.

Keywords: *Rhizobium*, phylogeny, nodulation gene, nitrogen fixation gene, common bean

INTRODUCTION

The legume-nodulating bacteria, collectively called rhizobia, form a group of soil bacteria belonging to the α-subclass of the *Proteobacteria* (Young & Haukka, 1996). The current taxonomy reveals their wide diversity at the genus, species and intraspecies levels. They are currently split into six genera: *Rhizobium*, *Sinorhizobium*, *Mesorhizobium*, *Bradyrhizobium*, *Azorhizobium* and the recently described genus *Allorhizobium* (de Lajudie *et al.*, 1998a). These genera are phylogenetically separate from each other based on 16S rDNA sequences as shown in Fig. 1, but the rhizobia do not form a coherent group since they are intermingled with other non-symbiotic bacteria (Young & Haukka, 1996). Three species whose phylogenetic position remains unclear might constitute two other potential genera: one represented by *Rhizobium galegae* (Lindström, 1989) and the closely related species *Rhizobium huautlense* (Wang *et al.*, 1998), and a second one by *Rhizobium giardinii* (Amarger *et al.*, 1997).

Rhizobial taxonomy does not globally reflect the symbiotic features of rhizobia, particularly their host plant range. A classification of rhizobia correlated with

†**Present address:** Agriculture and Agri-Food Canada, 1391 Sandford Street, London, Ont., Canada N5V 4T3.

‡**Present address:** Centre de Recherches et de Développement sur les Sols et les Grandes Cultures, Agriculture et Agroalimentaire Canada, 2560 Boul. Hochelaga, Sainte-Foy, Québec, Canada GIV 2J3.

Abbreviation: Sym genes, symbiotic genes.

The GenBank accession numbers for the sequences reported in this paper are AF217261 through AF217272 for *nodC* and AF218126, AF275670 and AF275671 for *nifH*.

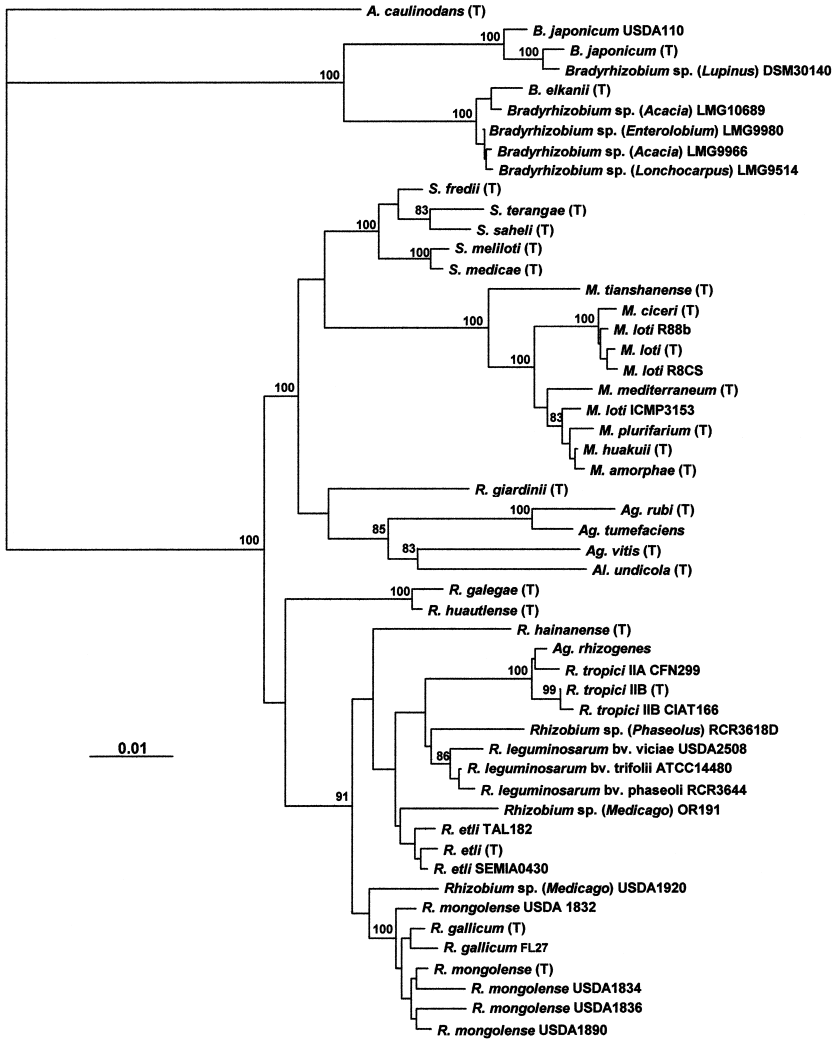

Fig. 1. Phylogenetic (neighbour-joining) tree showing the relationships between rhizobia and related agrobacteria based on 918 bp aligned 16S rDNA sequences (positions 20–937 of the nucleotide sequence of the *R. leguminosarum* bv. *trifolii* 16S rRNA gene). Significant bootstrap probability values are indicated at the branching points (only values greater than 80 % over 100 replicates are shown). This tree is almost identical to the tree based on 1450 bp aligned 16S rDNA sequences (not shown), which was constructed by omitting the 918 bp sequence of strain OR191. The main difference concerns the clade formed by *R. etli*, *R. leguminosarum* and *R. tropici*, which is no longer supported by significant bootstrap values in the tree including the sequence of strain OR191. The horizontal branches are drawn proportionally to the number of nucleotide substitutions per site. *A.*, *Azorhizobium*; *Ag.*, *Agrobacterium*; *Al.*, *Allorhizobium*; *B.*, *Bradyrhizobium*; *M.*, *Mesorhizobium*; *R.*, *Rhizobium*; *S.*, *Sinorhizobium*; (T), type strain.

symbiotic features collides with the complexity of the molecular mechanisms involved in host specificity (for review, see Perret *et al*., 2000), and also the difficulty of establishing host range (Pueppke & Broughton, 1999), particularly because of the great number of legume species. Such a classification requires a standardization of nodulation tests and the control of optimal conditions for plant growth. Although probably underestimated, it is well established that many rhizobia are able to nodulate different legume genera, and that many legumes can be nodulated by several rhizobial species. The bacterial nodulation (*nod*) genes, which are induced by plant flavonoids, determine the synthesis of Nod factors, the main nodulation signal molecules (for review, see Perret *et al*., 2000). Both the type and the amount of Nod factors are important in determining host specificity. However, rhizobia which have dissimilar *nod* genes and produce different Nod factors can effectively nodulate the same plant. For example, this is the case of *Rhizobium etli* bv. *phaseoli* and *Rhizobium tropici*, which are both nitrogen-fixing symbionts of *Phaseolus vulgaris*, the common bean plant (Poupot *et al*., 1993, 1995).

The convergence of different rhizobia that harbour distinct *nod* genes to the same nodulation phenotype has been reported by comparison of phylogenies inferred from housekeeping and Sym (symbiotic) gene loci (Haukka *et al*., 1998; Wernegreen & Riley, 1999). On the other hand, lateral gene transfer of the Sym genes appears to be the most plausible hypothesis to explain cases of phylogenetic incongruence between Sym and housekeeping genes (Martinez-Romero & Caballero-Mellado, 1996; Young & Haukka, 1996). Indeed, phylogenetic trees based on sequences of *nod* genes are generally not congruent with those based on 16S rDNA sequences, but the *nod* trees show some correlation with host plant range (Dobert *et al*., 1994; Lindström *et al*., 1995; Ueda *et al*., 1995; Haukka *et al*., 1998; Wernegreen & Riley, 1999). By contrast, the phylogeny of *nifH* genes, which encode the dinitrogenase reductase enzyme, has been reported as closely following that of 16S rRNA genes (Hennecke *et al*., 1985; Young, 1992; Dobert *et al*., 1994), despite some exceptions (Eardly *et al*., 1992). However, these *nifH* phylogenies were based on analysis of a small number of sequences. Recently, Haukka *et al*. (1998) analysed many more sequences and

concluded that, for rhizobia, the phylogeny of *nifH* was generally not consistent with the phylogeny of 16S rRNA, but was broadly similar to that of *nodA* genes. This result agrees with the fact that the *nod* and *nif* genes are often tightly linked in rhizobia, and can be located on transmissible elements such as plasmids in many rhizobial species or transposon-like elements in *Mesorhizobium loti* (Sullivan *et al*., 1995; Sullivan & Ronson, 1998).

Although it is widely agreed that phylogenies based on stable chromosomal genes are necessary to establish a biologically meaningful rhizobial taxonomy, a proper definition of broad host range should consider the diversity of the Sym genes rather than the diversity of the species that carry them. It would thus make sense to include the characterization and the phylogenetic classification of Sym genes in the minimal standards for the description of new rhizobia as previously proposed for the 16S rRNA gene sequences (Graham *et al*., 1991). Such a classification should provide a complementary basic framework for our understanding of the *Rhizobium*–legume symbiosis.

The *P. vulgaris* microsymbionts form a taxonomically heterogeneous group. So far, five recognized species plus two distinct 16S rDNA lineages have been described from rhizobial isolates recovered from bean nodules (Martinez-Romero *et al*., 1991; Segovia *et al*., 1993; van Berkum *et al*., 1996; Amarger *et al*., 1997; Herrera-Cervera *et al*., 1999). These species are distributed in two genera, *Rhizobium* and *Sinorhizobium*, and the potentially new genus represented by *R. giardinii*. Moreover, some of these species are subdivided in biovars based on the extent of host range and genetic characteristics of Sym genes (Amarger *et al*., 1997; Wang *et al*., 1999a). Our aim was to determine the congruence between classifications of rhizobia based on Sym genes and 16S rRNA and to estimate evolutionary relationships among rhizobia that have similar host legumes but are chromosomally diverse. Our main focus was the bean symbionts, but we have also analysed a larger collection of rhizobia. We initially developed a simple and rapid method to characterize Sym genes in rhizobia based on RFLP of PCR-amplified DNA, as previously achieved for 16S rRNA genes (Laguerre *et al*., 1994, 1997), and Sym gene loci in *Rhizobium leguminosarum* (Laguerre *et al*., 1996). Representative Sym genotypes were selected for subsequent phylogenetic analyses based on DNA sequencing. As a nodulation gene marker, we chose the *nodC* gene, which is a common *nod* gene essential for nodulation in all rhizobial species investigated so far. This gene encodes an *N*-acetylglucosaminyltransferase which is involved in the first step of Nod factor assembly, and it is also a determinant of host range (reviewed by Perret *et al*., 2000). In addition, the *nodC* sequence is relatively long, which enabled the PCR amplification of large DNA fragments to *a priori* ensure maximum specificity of RFLP fingerprints and maximum robustness of phylogeny inferred from nucleotide sequences. As a nitrogen fixation marker, we chose the *nifH* gene, for which the largest number of rhizobial sequences is available for comparison (Haukka *et al*., 1998).

METHODS

Bacteria. The 83 strains used in this study are listed in Table 1. The bacterial sample included reference or type strains representing 23 recognized species of *Rhizobium*, *Sinorhizobium*, *Mesorhizobium* and *Bradyrhizobium*, as well as some unclassified rhizobia from various host plants. The geographic origin and the source or the references were given for most of these rhizobia in previous studies dealing with their molecular characterization based on the analysis of the 16S rRNA genes (Laguerre *et al*., 1994, 1997). Three additional isolates from *P. vulgaris* were included. Strains HT2a2 and HT4c1 are from the collection of the Laboratoire de Microbiologie des Sols in Dijon, France. They were isolated from nodules of *P. vulgaris* (cv. Vernandon) grown in pots on soil collected at the INRA-Hagetmau experimental station located in the southwest of France. There was no known history of bean cultivation and inoculation at this site. The two strains have the same 16S rDNA type as *Rhizobium* sp. (*Medicago*) OR191 based on PCR-RFLP analysis (Laguerre *et al*., 1994, 1997; see Table 1). *Sinorhizobium* sp. (*Phaseolus*) GR-X8 was included for sequencing of *nifH* DNA. This strain is from the collection of the Estación Experimental del Zaidín in Granada, Spain. It was isolated from nodules of *P. vulgaris* (cv. Xera) and from the same soil as *Sinorhizobium* sp. GR-06 (Table 1). The two strains had the same 16S rDNA type, which was also identical to that of *Sinorhizobium fredii* strains (see Table 1).

PCR amplification and restriction digestion. Approximately 930 bp of the 1300 bp *nodC* gene was amplified by using forward primers nodCF, nodCFu, nodCF2, nodCF4 or nodCFn, and reverse primer nodCI (Table 2). These primers were designed by comparing available *nodC* sequences for *R. leguminosarum* bv. *viciae*, *R. tropici*, *R. etli* bv. *phaseoli*, *R. galegae*, *Sinorhizobium meliloti*, *S. fredii*, *Sinorhizobium* sp. NGR234, *M. loti*, *Bradyrhizobium japonicum* and *Bradyrhizobium elkanii*.

The primers used for PCR amplification of about 780 bp of the 890 bp *nifH* gene were designed by comparing known *nifH* sequences for *R. leguminosarum* bv. *trifolii*, *R. etli* bv. *phaseoli*, *S. meliloti*, *Sinorhizobium* sp. NGR234, *B. japonicum*, *Azorhizobium caulinodans*, *Azospirillum brasilense*, *Azotobacter chroococcum*, *Rhodobacter sphaeroides*, *Rhodobacter capsulatus* and *Rhodospirillum rubrum*.

The cell growth conditions and the target DNA preparation were as previously described (Laguerre *et al*., 1997). The *nodC* DNA was amplified from 5–10 µl lysed cell suspension mixed with all PCR reagents: polymerase reaction buffer (Gibco-BRL); 2·5 mM $MgCl_2$; 200 µM (each) dATP, dCTP, dTTP, dGTP; 0·4–0·8 µM (each) *nodC* primers; 0·04 U *Taq* DNA polymerase (Gibco-BRL) μl^{-1}. DNA amplification was done by using a standard temperature profile including an annealing temperature of 55 °C (Laguerre *et al*., 1994). The procedure for amplification of the *nifH* fragments was similar, except that reactions were made with 1·5 mM $MgCl_2$, 20 µM (each) dNTP and 0·1 µM (each) *nifH* primers and the annealing temperature was increased to 57 °C. For a few strains, multiple *nodC* bands were obtained. In these cases, a small piece of agarose containing the band of the expected size was aspirated by using a hypodermic needle and used as a template in a new PCR.

Restriction pattern analysis of the PCR products with the

Table 1. Strains used in this study and data of PCR-RFLP analysis of *nodC* and *nifH* gene fragments

The data are displayed as restriction patterns encoded by letters. Identical patterns are designated by the same letter(s) for each endonuclease; the on-line version of this paper (available at http://mic.sgmjournals.org) contains a supplementary table giving information on the size of the restriction fragments for each pattern. ND, Not determined; No, unsuccessful PCR amplification. T, Type strain of the species.

Strain	Species or 16S rDNA type*	Biovar†	Plant of origin	*nod* type‡	*nif* type‡	*nodC* restriction pattern						*nifH* restriction pattern			Source or reference§
						*Cfo*I	*Hae*III	*Hin*fI	*Msp*I	*Nde*II	*Rsa*I	*Hae*III	*Msp*I	*Nde*II	
USDA 2370T	*R. leguminosarum*	*viciae*	*Pisum sativum*	V 1	V 1	A	A	A	A	A	A	A	A	A	Laguerre *et al.* (1997)
JISP24	*R. leguminosarum*	*viciae*	*Pisum sativum*	V 2	V 2	A	A	A	B	A	A	B	B	B	Laguerre *et al.* (1994)
F5NP1E	*R. leguminosarum*	*viciae*	*Pisum sativum*	V 3	ND	B	B	A	A	A	B	ND	ND	ND	MSDJ
L113	*R. leguminosarum*	*viciae*	*Lens esculenta*	V 4	V 3	A	A	B	A	A	A	B	C	A	Laguerre *et al.* (1994)
8401	*R. leguminosarum*	*viciae*		V 5	V 4	C	C	A	A	B	C	C	B	C	Laguerre *et al.* (1994)
USDA 2071	*R. leguminosarum*	*trifolii*	*Trifolium repens*	T 1	T 1	D	D	C	C	C	D	D	D	D	Laguerre *et al.* (1997)
cc2480a	*R. leguminosarum*	*trifolii*	*Trifolium subterraneum*	T 2	No	D	E	D	D	D	E	No	No	No	Laguerre *et al.* (1994)
T143	*R. leguminosarum*	*trifolii*	*Trifolium pratense*	T 3	T 2	E	F	D	E	E	F	E	E	E	Laguerre *et al.* (1994)
H112	*R. leguminosarum*	*trifolii*	*Phaseolus vulgaris*	T 1	T 3	D	D	C	C	C	D	F	D	F	Laguerre *et al.* (1994)
H132	*R. leguminosarum*	*phaseoli*	*Phaseolus vulgaris*	P 1	P 1	F	G	E	F	F	G	G	F	G	Laguerre *et al.* (1997)
H251	*R. giardinii*	*phaseoli*	*Phaseolus vulgaris*	P 1	ND	F	G	E	F	F	G	ND	ND	ND	Laguerre *et al.* (1994)
RCR3618D	Type *Rhizobium* sp. RCR3618D	'*phaseoli*'	*Phaseolus vulgaris*	P 1	ND	F	G	E	ND	ND	G	ND	ND	ND	van Berkum *et al.* (1996)
VikingI	*R. etli*	*phaseoli*	*Phaseolus vulgaris*	P 2	P 1	F	G	E	G	F	G	G	F	G	Laguerre *et al.* (1994)
PhD12	*R. gallicum*	*phaseoli*	*Phaseolus vulgaris*	P 3	P 1	F	G	E	G	G	G	G	F	G	Laguerre *et al.* (1994)
H441	*R. leguminosarum*	*phaseoli*	*Phaseolus vulgaris*	P 4	P 1	G	G	F	H	H	H	G	F	G	Laguerre *et al.* (1994)
CFN 42T	*R. etli*	*phaseoli*	*Phaseolus vulgaris*	P 4	P 1	G	G	F	H	H	H	G	F	G	Laguerre *et al.* (1997)
CFN 299	*R. tropici*-IIA		*Phaseolus vulgaris*	Tr 1	Tr 1	H	H	B	I	I	I	H	G	H	Laguerre *et al.* (1997)
Airel 8, Airel 18	*R. tropici*-IIA		*Phaseolus vulgaris*	Tr 1	Tr 1	H	H	B	I	I	I	H	G	H	Amarger *et al.* (1994)
CIAT 899T	*R. tropici*-IIB		*Phaseolus vulgaris*	Tr 2	Tr 1	I	H	B	I	I	J	H	G	H	Laguerre *et al.* (1997)
BR859	*R. tropici*-IIB		*Phaseolus vulgaris*	Tr 2	Tr 1	I	H	B	I	I	J	H	G	H	Laguerre *et al.* (1994)
BR864	*R. tropici*		*Leucaena leucocephala*	Tr 2	Tr 2	I	H	B	I	I	J	H	G	I	Laguerre *et al.* (1994)
HT2a2, HT4c1	Type *Rhizobium* sp. OR191		*Phaseolus vulgaris*	Tr 1	ND	H	H	B	I	I	I	ND	ND	ND	MSDJ
R602spT, PhF29, PhP222	*R. gallicum*	*gallicum*	*Phaseolus vulgaris*	G 1	G 1	J	I	G	J	J	K	I	H	J	Laguerre *et al.* (1994); Amarger *et al.* (1997)
F127	*R. gallicum*	*gallicum*	*Phaseolus vulgaris*	G 2	G 2	J	J	H	K	K	L	J	H	K	Laguerre *et al.* (1994)
USDA 3497	Type *R. mongolense*	'*gallicum*'	*Leucaena leucocephala*	G 3	G 2	J	I	G	L	J	M	J	H	K	Laguerre *et al.* (1997)
H152T	*R. giardinii*	*giardinii*	*Phaseolus vulgaris*	Gi 1	No	K	K	H	M	L	N	No	No	No	Laguerre *et al.* (1994)
H241	*R. giardinii*	*giardinii*	*Phaseolus vulgaris*	Gi 1	ND	K	K	H	M	L	N	ND	ND	ND	Laguerre *et al.* (1994)
GR-06	Type *S. fredii*		*Phaseolus vulgaris*	Sp 1	Sp 1	L	L	I	N	M	O	K	I	L	Herrera-Cervera *et al.* (1999)
GR-X8	Type *S. fredii*		*Phaseolus vulgaris*	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	EEZ
USDA 1002T	*S. meliloti*		*Medicago sativa*	Me 1	Me 1	M	M	J	O	N	P	L	J	M	Laguerre *et al.* (1997)
RCR2011	*S. meliloti*		*Medicago sativa*	Me 2	Me 2	M	N	K	P	O	Q	L	K	M	Laguerre *et al.* (1994)
OR191	Type *Rhizobium* sp. OR191		*Medicago sativa*	Me 3	Me 3	N	O	L	P	P	R	M	L	N	Laguerre *et al.* (1997)
M1	*S. medicae*		*Medicago orbicularis*	Me 4	Me 4	O	P	M	P	N	S	N	M	O	Laguerre *et al.* (1997)
M102	*S. medicae*		*Medicago truncatula*	Me 4	Me 4	O	P	M	P	N	S	N	M	O	Laguerre *et al.* (1997)
USDA 1844T, USDA 1890	*R. mongolense*		*Medicago rustica*	Me 5	Me 5	P	O	L	P	Q	R	O	N	P	van Berkum *et al.* (1998)
NZP 2213T	*M. loti*		*Lotus tenuis*	Lo 1	Lo 1	Q	Q	N	Q	R	T	P	O	Q	Laguerre *et al.* (1997)
NZP 2037	*M. loti*		*Lotus divaricatus*	Lo 2	Lo 2	R	Q	O	Q	S	U	P	O	R	Laguerre *et al.* (1997)
NZP 2234	*M. loti*		*Lotus cornitulatus*	Lo 3	Lo 3	S	Q	M	R	T	T	Q	P	S	Laguerre *et al.* (1997)
MSDJ 865	*M. loti*		*Lotus cornitulatus*	Lo 4	Lo 4	S	Q	P	Q	S	T	P	P	S	Laguerre *et al.* (1997)
A-1BST	*M. tianshanense*		*Glycyrrhiza pallidiflora*	Ti 1	Ti 1	T	R	Q	S	U	V	R	Q	T	Chen *et al.* (1995)
UPM-Ca7T, USDA 3233	*M. ciceri*		*Cicer arietinum*	Ci 1	Ci 1	U	S	R	T	V	W	S	O	U	Laguerre *et al.* (1997)
UPMCa36T	*M. mediterraneum*		*Cicer arietinum*	Ci 2	Ci 1	U	S	S	T	W	W	S	O	U	Laguerre *et al.* (1997)
MSDJ 2184	Type *Mesorhizobium* sp. MSDJ 2184		*Coronilla varia*	Co 1	Co 1	V	T	T	U	X	X	ND	R	V	Laguerre *et al.* (1997)
7653R	*M. huakuii*		*Astragalus sinicus*	As 1	As 1	W	U	U	V	Y	Y	T	S	T	Yang *et al.* (1999)
Ra5	*M. huakuii*		*Astragalus sinicus*	As 1	As 2	ND	ND	U	V	ND	ND	U	S	T	Yang *et al.* (1999)
ACCC 19665T	*M. amorphae*		*Amorpha fruticosa*	Am 1	Am 1	X	V	E	W	Z	W	V	T	W	Wang *et al.* (1999b)
HL56, SH15003	*M. amorphae*		*Amorpha fruticosa*	ND	Am 1	ND	ND	ND	ND	ND	ND	V	T	W	Wang *et al.* (1999b)
ORS 1004, ORS 1010	*M. plurifarium*		*Acacia senegal*	MP 1	MP 1	Y	W	V	X	AA	Z	W	I	X	de Lajudie *et al.* (1998b)
ORS 1009T, ORS 1007	*S. terangae*	*acaciae*	*Acacia laeta*	Ac 1	Ac 1	Z	X	W	Y	AB	AA	X	U	Y	de Lajudie *et al.* (1994); Laguerre *et al.* (1997)
ORS 51	*S. terangae*	*sesbaniae*	*Sesbania rostrata*	No	Se 1	No	No	No	No	No	No	Y	V	Z	de Lajudie *et al.* (1994)
ORS 609T	*S. saheli*	*sesbaniae*	*Sesbania cannabina*	No	Se 2	No	No	No	No	No	No	Z	V	Z	Laguerre *et al.* (1997)
ORS 611	*S. saheli*	*sesbaniae*	*Sesbania grandiflora*	No	Se 2	No	No	No	No	No	No	Z	V	Z	de Lajudie *et al.* (1994)
So2T, S35	*R. huautlense*		*Sesbania herbacea*	Se 1	Se 3	AA	Y	X	Z	AC	AB	Y	V	AA	Wang *et al.* (1998)
HAMBI 540T	*R. galegae*		*Galega orientalis*	Ga 1	Ga 1	AB	Z	Y	AA	AD	AC	AA	W	AB	Laguerre *et al.* (1997)
USDA 205T	*S. fredii*		*Glycine max*	F 1	F 1	T	AA	Z	AB	AE	AD	AB	X	AC	Laguerre *et al.* (1997)
MSDJ 1536	*S. fredii*		*Glycine max*	F 1	F 1	T	AA	Z	AB	AE	AD	AB	X	AC	Laguerre *et al.* (1994)
MSDJ 1537	*S. fredii*		*Glycine max*	F 1	F 1	T	AA	Z	AB	AE	AD	AB	X	ND	Laguerre *et al.* (1994)

Table 1 (*cont.*)

Strain	Species or 16S rDNA type*	Biovar†	Plant of origin	*nod* type‡	*nif* type‡	*nodC* restriction pattern						*nifH* restriction pattern			Source or reference§
						*Cfo*I	*Hae*III	*Hin*fI	*Msp*I	*Nde*II	*Rsa*I	*Hae*III	*Msp*I	*Nde*II	
USDA 6^T, USDA 110	*B. japonicum*		*Glycine max*	J 1	J 1	AC	AB	AA	AC	AF	AE	AC	Y	AD	Laguerre *et al.* (1997)
USDA 123	*B. japonicum*		*Glycine max*	J 1	J 1	AC	AB	AA	AC	AF	AE	AC	Y	AD	USDA
G49	*B. japonicum*		*Glycine max*	J 1	J 1	AC	AB	AA	AC	AF	AE	AC	Y	AD	Lagacherie *et al.* (1997)
USDA 138	*B. japonicum*		*Glycine max*	J 1	J 1	AC	AB	AA	ND	ND	AE	AC	Y	AD	USDA
2281	*B. liaoningense*		*Glycine max*	J 1	J 1	AC	AB	AA	AC	AF	AE	AC	Y	AD	Xu *et al.* (1995)
USDA 76^T	*B. elkanii*		*Glycine max*	E 1	E 1	AD	AC	M	AD	AG	AE	AD	Z	AE	Laguerre *et al.* (1997)
USDA 61, USDA 94, USDA 340	*B. elkanii*		*Glycine max*	ND	E 1	ND	ND	ND	ND	ND	ND	AD	Z	AE	Kuykendall *et al.* (1992); Laguerre *et al.* (1997)
CB756	Type *Bradyrhizobium* sp. MSDJ 718		*Macrotyloma africanum*	Ma 1	Ma 1	AE	AB	M	AE	AH	AF	AE	AA	AF	Laguerre *et al.* (1994)
MSDJ 718	Type *Bradyrhizobium* sp. MSDJ 718		*Lupinus luteus*	Lu 1	Lu 1	AF	AD	AB	AF	AI	AG	AF	AB	AG	Laguerre *et al.* (1997)
LB84	*Bradyrhizobium* sp.		*Lupinus albus*	Lu 2	Lu 2	AG	AD	AC	AG	AJ	AG	AF	AB	AH	Lagacherie *et al.* (1983)
LB86	*Bradyrhizobium* sp.		*Lupinus albus*	ND	Lu 2	ND	ND	ND	ND	ND	ND	AF	AB	AH	Lagacherie *et al.* (1983)
VK4	Type *Bradyrhizobium* sp. VK4		*Lupinus* sp.	ND	Lu 3	ND	ND	ND	ND	ND	ND	AF	AC	AH	Bergersen *et al.* (1986)
WU425	Type *Bradyrhizobium* sp. VK4		*Ornithopus compressus*	Lu 3	Lu 3	AH	AE	AD	AH	AK	AH	AF	AC	AH	Bergersen *et al.* (1986)

* The 16S rDNA types were characterized by RFLP analysis of PCR-amplified 16S rRNA genes as previously reported (Laguerre *et al.*, 1994, 1997).

† The biovars indicated within quotation marks were identified in this study on the basis of the analysis of the Sym genes.

‡ The *nodC* and *nifH* types represent the combination of the restriction patterns; types sharing identical restriction patterns were given the same letter as follows: V, bv. *viciae*; T, bv. *trifolii*; P, bv. *phaseoli*; Tr, *tropici*; G, *gallicum*; Gi, *giardinii*; Me, *Medicago*; Lo, *Lotus*; Ti, *tianshanense*; Ci, *Cicer*; Co, *Coronilla*; As, *Astragalus*; Am, *amorphae*; Mp, *M. plurifarium*; Ac, bv. *acaciae*; Se, *Sesbania*; Ga, *Galega*; F, *fredii*; J, *japonicum*; E, *elkanii*; Ma, *Macrotyloma*; Lu, *Lupinus*. ND, Not determined; No, unsuccessful PCR amplification.

§ USDA, *Rhizobium* culture collection, Beltsville Agricultural Research Center, Beltsville, MD, USA; MSDJ, Laboratoire de Microbiologie des Sols, INRA, Dijon, France; EEZ, Estación Experimental del Zaidín, CSIC, Granada, Spain.

Table 2. Oligonucleotides used as PCR or sequencing primers

Primer	5′–3′ nucleotide sequence*	Position†
nodCF	AYGTHGTYGAYGACGGTTC	251–269
nodCFu	AYGTHGTYGAYGACGGITC	251–269
nodCF2	AYGTHGTYGAYGACGGCTC	251–269
nodCF4	AYGTHGTYGAYGACGGATC	251–269
nodCFn	AGGTGGTYGAYGACGGTTC	251–269
nodCI	CGYGACAGCCANTCKCTATTG	1160–1181
nifHF	TACGGNAARGGSGGNATCGGCAA	25–50
nifHI	AGCATGTCYTCSAGYTCNTCCA	787–808

* Y = C or T; H = A, C or T; R = A or G; S = C or G; K = G or T; N = A, C, G or T; I = inosine.

† Position of the primer nucleotide sequences in the corresponding *nodC* sequence of *R. leguminosarum* bv. *viciae* and *nifH* sequence of *R. leguminosarum* bv. *trifolii*.

restriction endonucleases listed in Table 1 was as previously described (Laguerre *et al.*, 1997).

Sequencing of *nodC* and *nifH* DNA. All the nucleotide sequences, apart from two *nifH* sequences (see below), obtained in this study were determined by Genome Express (Grenoble, France). In a first attempt, crude *nodC* PCR products were directly sequenced on both strands by using primers nodCF and nodCI. The less degenerate primer nodCFn, which matched the same oligonucleotide sequence as nodCF, was used for PCR amplification and sequencing of crude PCR products from strains PhD12 and H251. Sequencing of *nodC* DNA with nodCFn and nodCI was achieved for strains H132 and Viking I only after purification of PCR products by either 70% ethanol precipitation in the presence of 0·7 M ammonium acetate or (when multiple PCR products were obtained) extraction of the *nodC* fragment from agarose gels by using a QIAEX II gel extraction kit (Qiagen) followed by ethanol/ammonium acetate precipitation. The sequences of the *nifH* fragments and the *nodC* fragment of strains R602sp, H152, ACCC 19665, USDA 2071 and OR191 were obtained after cloning the PCR products by using either a pT7Blue T-Vector kit (Novagen), a pGEM-T Vector kit (Promega) or a TOPO TA cloning kit (Invitrogen) according to the manufacturers' instructions. The result of cloning was checked by PCR amplification by using the vector plasmid primers T7 and SP6 according to the procedure described by Novagen. The crude PCR products were directly used for sequencing with both primers T7 and SP6. Sequencing of *nifH* fragments of strains GR-06 and GR-X8 was done in an ABI 373 XL Stretch Sequencer (Perkin-Elmer Biosystems) using an ABI Prism BigDye Terminator kit and vector-based sequencing primers (M13 universal and reverse primers).

From all these experiments, 450–750 nucleotides of each DNA strand were determined and an 808–935 bp sequence of *nodC* fragments and a 736–783 bp sequence of the *nifH* gene were reconstituted for each strain. Restriction site analyses of the sequences were performed by using the Bisance software (Dessen *et al*., 1990).

Phylogenetic analysis. The sequences have been deposited in the GenBank database under accession numbers AF217261 through AF217272 for the *nodC* sequences and AF218126, AF275670 and AF275671 for the *nifH* sequences of strains R602sp, GR-X8 and GR-06, respectively.

The accession numbers of the published sequences used for comparisons were as follows. The numbers for the *nodC* sequences were: M13658 (*R. leguminosarum* bv. *viciae* 238), X98514 (*R. tropici* IIA CFN 299), X87578 (*R. galegae* HAMBI 1174), M11268 (*S. meliloti* 1021), M73699 (*S. fredii* USDA 257), X73362 (*Sinorhizobium* sp. NGR234), X52958 (*M. loti* NZP 2037), U53327 [*Mesorhizobium* sp. (Oxytropis) N33], AF105431 (*Bradyrhizobium* sp. SNU001), L18897 (*Azorhizobium caulinodans* ORS 571^T). The numbers for the *nifH* sequences were: K00490 (*R. leguminosarum* bv. *trifolii* SU329), M55225 (*R. tropici* IIB CIAT 899^T), M15942 (*R. etli* bv. *phaseoli* CFN 42^T), M55227 (*R. etli* bv. *phaseoli* Olivia-4), AF107621 (*R. etli* bv. *mimosae* Mim2), M55226 (*Rhizobium gallicum* bv. *gallicum* FL27), M55228 [*Rhizobium* sp. (*Medicago*) OR191], Z95230 [*Rhizobium* sp. (*Lonchocarpus*) BR6001], J01781 (*S. meliloti* 41), M55232 (*S. meliloti* USDA 1002^T), M55232 (*S. meliloti* CC2013), M55231 (*Sinorhizobium medicae* CC169), Z95229 (*S. fredii* USDA 191), Z95218 (*Sinorhizobium terangae* bv. *acaciae* ORS 1009^T), SSZ95221 (*Sinorhizobium saheli* bv. *sesbaniae* ORS 609^T), AE000105 (*Sinorhizobium* sp. NGR234), Z95224 [*Sinorhizobium* sp. (*Acacia*) HAMBI 1499], Z95212 [*Sinorhizobium* sp. (*Leucaena*) BR827], Z95213 [*Sinorhizobium* sp. (*Prosopis*) M6], Z95228 [*Mesorhizobium* sp. (*Leucaena*) INPA78B], K01620 (*B. japonicum* USDA 110), M16709 (*Azorhizobium caulinodans* ORS 571^T), X51500 (*Azospirillum brasilense* Sp7), X03916 (*Azotobacter chroococcum* MCD1), X07866 (*Rhodobacter capsulatus* SB1003), J01740 (*Klebsiella pneumoniae*). The numbers for the 16S rDNA sequences were: U89831 (*R. leguminosarum* bv. *viciae* USDA 2508), X67227 (*R. leguminosarum* bv. *trifolii* ATCC 14480), U29388 (*R. leguminosarum* bv. *phaseoli* RCR3644), X67233 (*R. tropici* IIA CFN 299), X67234 (*R. tropici* IIB CIAT 899^T), U38469 (*R. tropici* IIB CIAT 166), U28916 (*R. etli* CFN 42^T), U28939 (*R. etli* TAL182), U47303 (*R. etli* SEMIA0430), U86343 (*R. gallicum* R602sp^T), AF008129 (*R. gallicum* FL27), U89817 (*Rhizobium mongolense* USDA 1844^T), U89816 (*R. mongolense* USDA 1832), U89818 (*R. mongolense* USDA 1834), U89819 (*R. mongolense* USDA 1836), U89821 (*R. mongolense* USDA 1890), U89823 [*Rhizobium* sp. (*Medicago*) USDA 1920], X91211 [*Rhizobium* sp. (*Medicago*) OR191], U29387 [*Rhizobium* sp. (*Phaseolus*) RCR3618D], U71078 (*Rhizobium hainanense* I66^T), D12793 (*R. galegae* HAMBI 540^T), AF025852 (*R. huautlense* S02^T), U86344 (*R. giardinii* H152^T), D12783 (*S. meliloti* USDA 1002^T), D12783 (*S. medicae* A321^T), X67231 (*S. fredii* USDA 205^T), X68388 (*S. terangae* ORS 1009^T), X68390 (*S. saheli* ORS 609^T), X67229 (*M. loti* NZP 2213^T), U50164 (*M. loti* R8CS), U50165 (*M. loti* R88b), U50166 (*M. loti* ICMP 3153), D12797 (*Mesorhizobium huakuii* CCBAU 2609^T), U07934 (*Mesorhizobium ciceri* UPM-Ca7^T), L38825 (*Mesorhizobium mediterraneum* UPM-Ca36^T), U71079 (*Mesorhizobium tianshanense* A-1BS^T), Y14158 (*Mesorhizobium plurifarium* LMG 11892^T), AF041442 (*Mesorhizobium amorphae* ACCC 19665^T), U69638 (*B. japonicum* USDA 6^T), Z35330 (*B. japonicum* USDA 110), U35000 (*B. elkanii* USDA 76^T), X87273 [*Bradyrhizobium* sp. (*Lupinus*) DSM 30140], X70405 [*Bradyrhizobium* sp. (*Acacia*) LMG 10689], X70403 [*Bradyrhizobium* sp. (*Acacia*) LMG 9966], X70404 [*Bradyrhizobium* sp. (*Enterolobium*) LMG 9980], X70401 [*Bradyrhizobium* sp. (*Lonchocarpus*) LMG 9514], Y17047 (*Allorhizobium undicola* LMG 11875^T), X67221 (*Azorhizobium caulinodans* ORS 571^T), X67223 (*Agrobacterium tumefaciens* LMG 196), X67228 (*Agrobacterium rubi* LMG 156^T), X67225 (*Agrobacterium vitis* LMG 8750^T), X67224 (*Agrobacterium rhizogenes* LMG 152).

Molecular sequence analyses were performed by using programs available in the Bisance software. Nucleotide and amino acid sequences were aligned with CLUSTAL W (Thompson *et al*., 1994). Phylogenetic trees of *nodC*, *nifH* and 16S rRNA genes were inferred by using the Phylogenetic Inference Package (PHYLIP; Felsenstein, 1989) with neighbour-joining analyses from Kimura's (Kimura, 1980) two-parameter nucleotide distances, and the maximum-likelihood method. Phylogenetic trees of NodC and NifH proteins were constructed using the neighbour-joining method from Dayhoff PAM distance matrix computed with the PROTDIST program of PHYLIP. Confidence in neighbour-joining trees was assessed by bootstrap analysis with the SEQBOOT and CONSENSE programs of PHYLIP.

RESULTS AND DISCUSSION

PCR-amplification of *nodC* and *nifH* gene fragments

The *nodC* gene fragment was amplified for the majority of the strains investigated (67/82) by using primer pair nodCF–nodCI. Amplification was achieved for *R. galegae* HAMBI 540 by using either nodCF2 or nodCFu, and for *B. elkanii* USDA 76 by using nodCF4, instead of nodCF. Forward primer nodCFu was used with *Bradyrhizobium* sp. LB84 and WU425, and *M. amorphae* ACCC 19665. The *nodC* gene could not be amplified for the strains of the bv. *sesbaniae* of *S. terangae* and *S. saheli* by using the *nodC* primers described in this study, while PCR amplification was successful for the strains of *S. saheli* and *S. terangae* bv. *acaciae*. The bv. *sesbaniae* has been created recently to group *S. saheli* and *S. terangae* strains isolated from *Sesbania* spp. that showed similar genetic and phenotypic symbiotic features (Lortet *et al*., 1996; Haukka *et al*., 1998). The absence of an amplification product was probably due to some nucleotide mismatches in *nodC* genes of bv. *sesbaniae* strains with the 3′ end region of one or both oligonucleotides used as *nodC* primer pairs. Amplification of the *nifH* fragment was achieved for all the 76 strains investigated, except for *R. leguminosarum* bv. *trifolii* cc2480a for unknown reasons, and for *R. giardinii* bv. *giardinii* H152 as expected since strains in this biovar probably lack *nifKDH* genes (Geniaux *et al*., 1993; Amarger *et al*., 1997).

Correlation between RFLP of Sym genes and host specificity

The results of the RFLP analysis of the PCR-amplified *nodC* and *nifH* fragments are given in Table 1. From the combined data with each gene, we identified 45 composite *nod* types and 41 *nif* types among the 82 and 76 strains investigated, respectively. The combination of

the *nod* and *nif* gene analyses revealed 50 symbiotic (*nod–nif*) genotypes. We concluded that both genes were highly polymorphic among species and biovars.

Intraspecies polymorphism was also detected among strains isolated from the same host legume (in *R. tropici*, *S. meliloti*, *M. loti* and *M. huakuii*) and, irrespective of the species, within previously defined biovars (*viciae*, *trifolii*, *phaseoli*, *gallicum*, *sesbaniae*), but differences between pairs of restriction patterns could be simply explained in terms of gain or loss of only one or two restriction sites. These results indicate that the genes are closely related.

Unclassified strains could be assigned to previously defined biovars. Thus *Rhizobium* sp. (*Phaseolus*) RCR3618D had Sym genes characteristic of bv. *phaseoli*. *Rhizobium* sp. (*Leucaena*) USDA 3497 could be classified into bv. *gallicum*, which is consistent with the ability of *R. gallicum* bv. *gallicum* to nodulate *Leucaena leucocephala* (Amarger *et al.*, 1997).

More generally and independent of their taxonomic status, the strains sharing at least two similar *nodC* and/or *nifH* restriction patterns originated from host plants belonging to the same species or genus, or to the same known cross-inoculation group. The *Rhizobium* sp. (*Phaseolus*) strains HT2a2 and HT4c1 had *nodC* genes typical of *R. tropici*. The soybean bradyrhizobia *B. japonicum* and *Bradyrhizobium liaoningense* had identical *nod* and *nif* types. Most restriction patterns were similar in the chickpea mesorhizobia *M. ciceri* and *M. mediterraneum*. The strains of *R. mongolense* and *Rhizobium* sp. OR191 which originated from *Medicago* species shared four *nodC* restriction patterns and the ability to form nitrogen-fixing nodules with common beans (Eardly *et al.*, 1992; van Berkum *et al.*, 1998). However, only the latter was reported to nodulate *Leucaena leucocephala* (Del Papa *et al.*, 1999). The *Medicago* sinorhizobia *S. meliloti* and *S. medicae* also showed two identical *nodC* restriction patterns, which indicates that they have closely related *nodC* genes. Similar *nifH* restriction patterns were obtained among the *R. huautlense* strains isolated from *Sesbania herbacea* and the bv. *sesbaniae* of *S. terangae* and *S. saheli*, showing that at least their *nifH* genes are closely related. However, the *nodC* genes of the *R. huautlense* strains were easily amplified by contrast to those of the *Sesbania* sinorhizobia. This result indicated some nucleotide differences between their *nodC* genes but further investigations are needed to estimate to what extent their nodulation genes differ. The five unclassified bradyrhizobia isolated from *Lupinus* or *Ornithopus* species, two legume genera that form a single cross-inoculation group (Graham, 1976), also had closely related *nifH* genes.

Conversely, strains with no known common host plants shared few restriction patterns (and no more than one in pairwise comparisons). However, there were also strains among *Phaseolus*, *Medicago*, *Acacia* or soybean rhizobia that did not share more than one restriction pattern. The RFLP method was not suitable for obtaining more information about phylogenetic relationships between the genes. Few restriction sites in *nodC* and *nifH* gene sequences were actually conserved among species or biovars, according to the available nucleotide sequences. Therefore, it was not possible to map the restriction sites for rigorous phylogenetic analyses, and so further nucleotide sequencing was required.

Relationship of phylogeny of Sym genes to host specificity

Complete or partial sequences of the PCR-amplified *nodC* fragments were determined for representatives of the different RFLP *nod–nif* types among strains that originated from *Phaseolus vulgaris* nodules, and for *R. leguminosarum* bv. *trifolii* M37, *Rhizobium* sp. OR191 from *Medicago sativa*, and the type strain ACCC 19665 of the recently described species *M. amorphae* (Wang *et al.*, 1999b). We also determined the sequence of the *nifH* fragment of *R. gallicum* bv. *gallicum* R602sp and of the *Sinorhizobium* sp. (*Phaseolus*) strains GR-06 and GR-X8. Restriction site mapping and comparisons with the experimental RFLP data confirmed the quality of the sequences. Phylogenetic analysis of *nodC* and *nifH* sequences was performed by using the neighbour-joining and the maximum-likelihood methods, which led to similar results, except for some uncertain nodes which were not supported by high bootstrap values ($>85\%$ over 500 replicates) in the *nifH* neighbour-joining tree.

The *nodC* trees were also similar to that derived from protein translation of the DNA sequences and only the neighbour-joining tree is shown in Fig. 2. The *nodC* phylogeny was well correlated with the host plant range. All the *nodC* genes but one of the *Phaseolus* rhizobia formed a robust cluster within which the similarity values ranged from 81·9 to 99·5%. This result suggests that these *nod* genes evolved from a common ancestor. *R. tropici* fell outside the bean symbiont cluster and the present data did not show evidence that *R. tropici* was more strongly associated with the bean symbionts than with other rhizobia. Each of the bvs *phaseoli*, *gallicum* and *giardinii* corresponded to distinct lineages or subclusters, which correlates with the differences observed in host plant range between these biovars (Amarger *et al.*, 1997; Sessitsch *et al.*, 1997). However, a close relationship was found between the *nodC* genes of *R. gallicum* bv. *gallicum* and *R. giardinii* bv. *giardinii*, a result that could not have been anticipated from the RFLP analysis. The similarity values were higher than 90%. Strain GR-06 formed its own lineage within the bean symbiont cluster, and consequently this strain could not be assigned to any of the previously defined biovars.

The *nodC* gene of *Rhizobium* sp. (*Medicago*) OR191 clustered with the *S. meliloti nodC* gene, its closest relative (87·8% homology), suggesting that these genes originated from a common ancestor. OR191 was reported to form ineffective nodules with its host of origin, *Medicago sativa*, but to be symbiotically more effective with common bean (Eardly *et al.*, 1992).

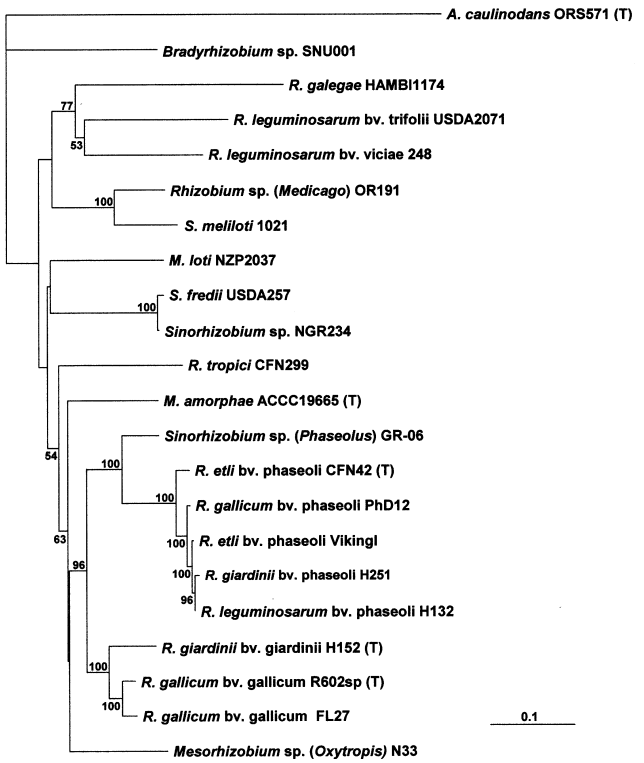

Fig. 2. Phylogenetic (neighbour-joining) tree based on 781 bp aligned sequences of *nodC* (positions 337–1113 of the nucleotide sequence of the *R. leguminosarum* bv. *viciae nodC* gene). Only bootstrap probability values greater than 50% (over 100 replicates) are indicated at the branching points. The horizontal branches are drawn proportionally to the number of nucleotide substitutions per site. *A.*, *Azorhizobium*; *B.*, *Bradyrhizobium*; *M.*, *Mesorhizobium*; *R.*, *Rhizobium*; *S.*, *Sinorhizobium*; (T), type strain.

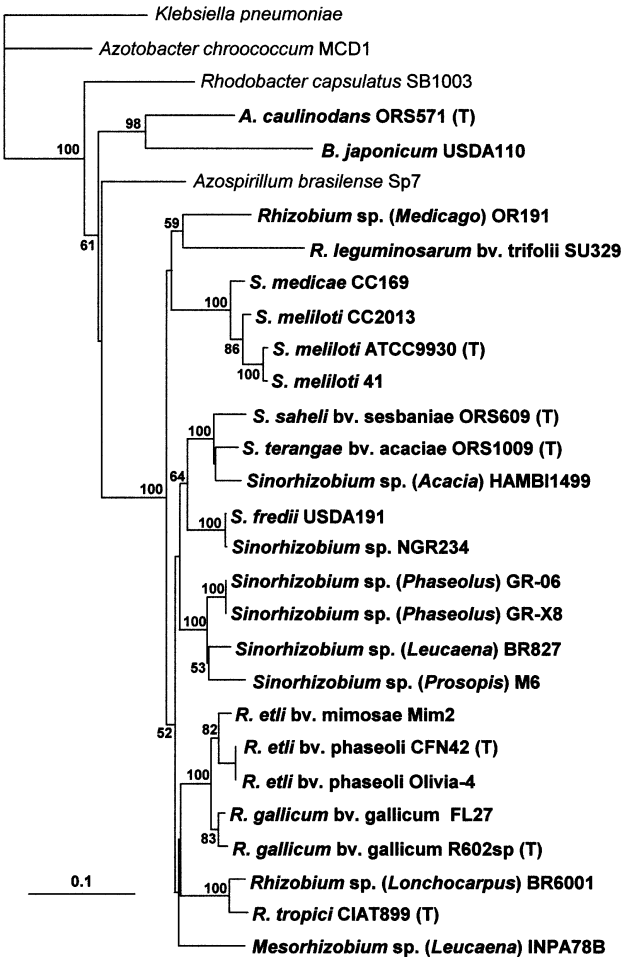

Fig. 3. Phylogenetic (neighbour-joining) tree based on 475 bp aligned sequences of *nifH* (positions 312–786 of the nucleotide sequence of the *R. leguminosarum* bv. *trifolii nifH* gene). Only bootstrap probability values greater than 50% (over 500 replicates) are indicated at the branching points. The horizontal branches are drawn proportionally to the number of nucleotide substitutions per site. *A.*, *Azorhizobium*; *B.*, *Bradyrhizobium*; *M.*, *Mesorhizobium*; *R.*, *Rhizobium*; *S.*, *Sinorhizobium*; (T), type strain.

However, phylogenetic analysis clearly showed that the *nodC* gene of OR191 was relatively distant from those of the *Phaseolus* symbionts.

The *nifH* neighbour-joining tree is shown in Fig. 3. A tree derived from the protein translation was also constructed (not shown), but the phylogenetic analysis was not reliable since the bootstrap values were generally low (<50%). Analysis of longer nucleotide sequences of *nifH* (738 bp aligned) improved the general robustness of *nifH* phylogenetic trees (not shown), but the number of rhizobial sequences available for comparison was too small for phylogenetic use. Our neighbour-joining tree was similar to those reported previously (Haukka *et al*., 1998; Wang *et al*., 1999a). As in the *nodC* tree, *R. gallicum* bv. *gallicum* formed a tight cluster with *R. etli* bv. *phaseoli*, which also included *R. etli* bv. *mimosae* Mim2. This strain was isolated from *Mimosa affinis* but is able to nodulate *P. vulgaris* and *Leucaena leucocephala* like strains of *R. gallicum* bv. *gallicum* (Wang *et al*., 1999a). On the basis of the lack of polymorphism revealed by the RFLP analysis (Table 1), it seems probable that the *nifH* genes are very similar among all the strains classified in bv. *phaseoli*, whatever the species. By contrast, but consistent with the RFLP data in *R. gallicum* bv. *gallicum*, the *nifH* nucleotide sequence of R602sp, the type strain of the species isolated in France, differed from that of the Mexican strain FL27 at 2% of nucleotide sites. This intra-biovar difference was relatively high compared to the differences observed between R602sp and *R. etli* bvs *mimosae* and *phaseoli* strains, which were only of 3·3 and 3·5%, respectively. The grouping of the two *R. gallicum* bv. *gallicum* strains into a subcluster was supported by an 83% bootstrap value, but not corroborated by the maximum-likelihood tree, in which the two *R. gallicum* strains and *R. etli* bv. *mimosae* Mim2 each formed a distinct lineage within the *Phaseolus* cluster. So far, almost all isolates of *R. gallicum* have been obtained from Europe (Amarger *et al*., 1997; Sessitsch *et al*., 1997; Herrera-Cervera *et al*., 1999). Because of the Mesoamerican origin of bean and

of *R. gallicum* strain FL27, Sessitsch *et al.* (1997) suggested that *R. gallicum* might have been imported to Europe as a seed contaminant. However, taking into account the relative divergence among the Sym genes of the French and Mexican *R. gallicum* bv. *gallicum* strains, this hypothesis appears unlikely.

The *R. tropici* sequence was positioned inside a clade including the *R. etli*–*R. gallicum* cluster (Fig. 3). The result was supported by the maximum-likelihood analysis, but not corroborated by the bootstrap analysis. The *Phaseolus* sinorhizobia GR-06 and GR-X8 have similar *nifH* gene sequences which showed a significant close relationship only with two tropical sinorhizobia isolated from *Leucaena* and *Prosopis* spp. trees (Haukka *et al.*, 1998). The relatedness of this cluster to the *R. gallicum*–*R. etli* cluster could not be resolved. In the neighbour-joining tree, the cluster including the *Phaseolus* sinorhizobia fell outside the *R. gallicum*–*R. etli* cluster, while in the maximum-likelihood tree, the two clusters grouped in a clade. However, the *nodA* genes of the tree sinorhizobia BR827 and M6 did not appear to be closely related to that of *R. etli* (Haukka *et al.*, 1998). Although we lack extensive data to assume that *nodA* and *nodC* phylogenies are equivalent, this seems likely based on previous phylogenetic surveys (Lindström *et al.*, 1995; Ueda *et al.*, 1995; Zhang *et al.*, 2000). No case of incongruence was detected by comparing our *nodC* tree with the *nodA* tree from Haukka *et al.* (1998), and congruence between *nodB* and *nodC* phylogenies, at least within the *Rhizobium* and *Sinorhizobium* genera, was also reported by Wernegreen & Riley (1999). If so, that would mean that *nod* genes of GR-06 are not closely related to those of the sinorhizobia BR827 and M6 and, consequently, that *nod* and *nif* genes did not co-evolve together in GR-06-like strains. The *nod* genes may be under a higher selective pressure to adapt to host differences, or alternatively, the *nod* and *nif* genes were not acquired together.

The second major difference between the *nodC* and *nifH* trees concerned the *Medicago* strain OR191, which was not associated with the tight cluster formed by *S. meliloti* and *S. medicae* in the *nifH* trees. The *nifH* sequence of OR191 formed its own lineage. Our updated analysis confirmed the earlier results of Eardly *et al.* (1992) indicating that this sequence was substantially different from those of the other rhizobia (similarity values lower than 89%). Again, the data suggest that the *nodC* and the *nifH* genes did not co-evolve together in this strain. The *nodC* gene of OR191 could be the ancestor from which *Medicago* strain genes diverged. Alternatively, OR191, and also GR-06-like strains, may have acquired only the *nod* genes by horizontal transfer from host-specific ancestral *nod* genes. Since *nod* and *nif* genes are often closely linked, another hypothesis would be that additional events of gene exchange and internal genetic rearrangements might have followed the co-transfer of the *nod* and *nif* genes. High-frequency rearrangements in plasmids of rhizobial strains involving recombination among reiterated sequences have been reported (see Garcia-de los Santos *et al.*, 1996; Romero *et al.*, 1998). Previous reports also indicated that some rhizobial strains, such as *S. saheli*, had different phylogenetic locations within rhizobia by comparing common *nod* gene and *nifH* evolutionary trees (Dobert *et al.*, 1994; Haukka *et al.*, 1998). Clearly, more work is needed to assess the linkage of *nod* and *nif* genes and the rest of the symbiotic genome.

Incongruence between symbiotic types and 16S-rDNA-based classification

Our results extend previously reported evidence showing that distinct rhizobial species can share similar Sym genes, and, conversely, that distinct Sym genotypes and phenotypes can be harboured by similar genomic backgrounds as defined by 16S rDNA types. Also, we observed additional cases of incongruence between the classifications and phylogenies resulting from comparative analyses of Sym genes and 16S rRNA (Figs 1–3).

The 16S rDNA type of *Rhizobium* sp. bv. *gallicum* USDA 3497 matched that of the *R. mongolense* type strain (Table 1), though it should be noted that this type is closely related to that of *R. gallicum* (Laguerre *et al.*, 1997; van Berkum *et al.*, 1998; Fig. 1). On the other hand, the *nodC* gene of *R. mongolense* was found to be close to that of the *Medicago* rhizobium OR191 (Table 1) and then probably distant from those of bv. *gallicum* on the basis of sequence analysis of the OR191 *nodC* gene. Although *R. mongolense* strains formed nitrogen-fixing nodules on *P. vulgaris* (van Berkum *et al.*, 1998), they were not able to nodulate *Leucaena leucocephala* in contrast with strains in bv. *gallicum*.

Strains HT2a2 and HT4c1 harboured *nodC* genes typical of *R. tropici*, but their 16S rDNA type was identical to that of strain OR191 (Table 1). This 16S rDNA type differed in nine restriction sites from that of the *R. tropici* type strain (Laguerre *et al.*, 1997), and the sequence data indicated that strain OR191 does not show significant phylogenetic affinity for *R. tropici* or any other rhizobial lineage (Fig. 1). Strains OR191, HT2a2 and HT4c1 may form a new *Rhizobium* species that would harbour different Sym genes.

The classification resulting from the Sym gene analysis fully reflected the host specificity for the sample of mesorhizobia investigated, while the 16S-rDNA-based phylogeny was irrespective of the host plant (Sullivan *et al.*, 1996; Laguerre *et al.*, 1997; de Lajudie *et al.*, 1998b; Fig. 1). In particular, the sample of four strains of *M. loti* examined represented three distinct 16S rDNA types intermixed with other mesorhizobia from various host legumes (Laguerre *et al.*, 1994, 1997), but they had clearly closely related *nodC* and *nifH* genes.

Similarly, the samples of soybean and lupin bradyrhizobial strains investigated in this study constituted heterogeneous 16S rDNA groups that were phylogenetically intermixed with each other and with bradyrhizobia isolated from other host legumes (Laguerre *et al.*, 1994, 1997). The *nodC* gene was polymorphic among the lupin bradyrhizobia, but the *nifH* gene sequences appear to be conserved within this group (Table 1). The

soybean species *B. japonicum* and *B. liaoningense* that were delineated by 16S rDNA sequence comparison (Xu *et al*., 1995) harboured similar Sym genes. The *B. elkanii* strains had a specific Sym RFLP type, but former phylogenetic analyses had established that the common nodulation genes of *B. japonicum* and *B. elkanii* were closely related (Dobert *et al*., 1994; Ueda *et al*., 1995). However, further studies are necessary to investigate whether the Sym gene phylogeny is correlated with the 16S rRNA classification and host range within the *Bradyrhizobium* genus, which includes a wide variety of yet unclassified microsymbionts associated with many legumes.

The *nod* genes of the soybean symbionts *S. fredii* and *B. japonicum* have been reported as being of the same lineage (Dobert *et al*., 1994), but the result was not supported by bootstrap analysis, as confirmed by more recent phylogenetic surveys inferred from *nodA* (Haukka *et al*., 1998) and *nodC* (Prévost *et al*., 2000) genes. Also, the *nifH* genes were not found to be related in these two species (Dobert *et al*., 1994; Haukka *et al*., 1998; Fig. 3). Therefore, among the soybean symbionts, the Sym gene phylogenies agree with the 16S rRNA phylogeny at the genus level.

By contrast, the *nodC* phylogeny was not only irrespective of the classification into species but also of the classification into genera among the *Phaseolus* rhizobia. Indeed, the highest similarity values (99–99·5%) between the *nodC* genes of the *Phaseolus* rhizobia were found between species rather than within species (only 96·5% similarity between the two *R. etli* bv. *phaseoli* strains, and a maximum of 84·6% between biovars within *R. gallicum* and *R. giardinii*). Almost all the *Phaseolus* symbionts belong to the genus *Rhizobium*, but GR-06-like strains were classified into the genus *Sinorhizobium*. The 16S rDNA sequence of strain GR-06 was found to be identical to that of *S. fredii* (Herrera-Cervera *et al*., 1999), but their Sym genes are not closely related (Figs 2 and 3). These results are consistent with the observation that GR-06-like strains were not able to nodulate soybean (Herrera-Cervera *et al*., 1999). *R. giardinii* provides further evidence of incongruence between *nodC* and 16S-rDNA-based phylogenies, the latter indicating that *R. giardinii* would deserve a genus status distinct from the genera described so far (Amarger *et al*., 1997), as confirmed by our updated phylogenetic tree (Fig. 1).

Additional evidence of discrepancy between Sym gene phylogeny and 16S-rDNA-based classification of rhizobia in genera was obtained within the *Medicago* and the *Sesbania* symbionts (Table 1; Figs 1 and 2). A similar case was reported by Wernegreen & Riley (1999) for *Rhizobium* sp. strains isolated from *Glycyrrhiza* and *Hedysarum* species that had *nod* genes closely related to those of *Medicago* sinorhizobia.

Specificity of the rhizobia–*P. vulgaris* associations

The *Phaseolus* rhizobia investigated in this study constitute a representative sample of the *Phaseolus* symbionts described so far. For all the strains except *R. tropici*, the ability to establish a symbiosis in beans appears to be directed by a specific set of diversified but closely related nodulation genes. This should lead to a reconsideration of the usually accepted view that *P. vulgaris* is an undiscriminating host based on the diversity of its microsymbionts and the fact that most rhizobia studied so far are able to nodulate *P. vulgaris* when tested in laboratory assays (Martinez *et al*., 1985; Michiels *et al*., 1998). Moulin *et al*. (2000) indicated that *nodA* phylogeny gives indications on structural features of Nod factors. It is possible that the bean symbionts that have closely related *nod* genes produce similar Nod factors adapted for specific bean receptors. These Sym genes may confer upon the bacteria that harbour them a good competitive ability for nodule formation on beans. Furthermore, bean isolates that were effective in nitrogen fixation have been found in each species and biovar in which they are included (Young, 1985; Martinez-Romero *et al*., 1991; Segovia *et al*., 1993; Amarger *et al*., 1994, 1997; Wang *et al*., 1999a), except in *R. giardinii* (Amarger *et al*., 1997). By contrast, the other rhizobia capable of nodulating bean plants as single-strain inoculants generally produce ineffective nodules (Martinez *et al*., 1985; Michiels *et al*., 1998). Taken together, these data suggest a certain specificity in the rhizobia–bean symbiosis.

Gene transfer is probably involved in evolution of the symbiotic functions

Assuming that *R. etli* bv. *phaseoli* strains would be the symbionts that co-evolved with *P. vulgaris*, it has been hypothesized that they may be the original donors of the Sym plasmid in *R. leguminosarum* bv. *phaseoli* (Segovia *et al*., 1993). Likewise, *R. gallicum* and *R. giardinii* bv. *phaseoli* probably received their Sym plasmids from *R. etli* bv. *phaseoli* or, more plausibly, from the *R. leguminosarum* bv. *phaseoli* strains which co-existed in the fields in which the *R. gallicum* and *R. giardinii* isolates originated (Amarger *et al*., 1997). The hypothesis of interspecies gene transfer is supported by the high similarity of the Sym genes among the bv. *phaseoli* subgroups and the co-occurrence of all these species in Europe (Geniaux *et al*., 1993; Sessitsch *et al*., 1997; Herrera-Cervera *et al*., 1999). In the same way, the finding that the Sym genes of strains HT2a2 and HT4c1 and of *R. tropici* are similar although their 16S rDNA types are relatively distant, and the co-occurrence of these rhizobia in the southwest of France (Amarger *et al*., 1994), argue for interspecies gene transfer. Hence, these data support the view that gene transfer would play a role in diversification and in structuring the natural populations of rhizobia, notably those nodulating *P. vulgaris*.

Furthermore, this work has also revealed the probable common origin of *nod* and *nif* genes among rhizobia belonging to the *Rhizobium* and *Sinorhizobium* genera that nodulate *Phaseolus*, *Medicago* and *Sesbania*. These findings suggest gene exchange events across genera,

which does not support the view of Wernegreen & Riley (1999) that the Sym genes in rhizobial genera diverge independently. Additionally, analysis of glutamine synthetase genes also suggests transfers of genes between the rhizobial genera (Turner & Young, 2000).

The comparison of the *nod*, *nif* and 16S rRNA phylogenies and the substantial correlation that we found between symbiotic genotypes and host plant groups taken together support the generally accepted hypothesis that lateral transfer of Sym genes and genetic rearrangements are involved in the acquisition and evolution of rhizobial symbiotic functions.

ACKNOWLEDGEMENTS

We thank M. R. Allard, M. Bours, F. Maillard, F. Revoy, S. Munoz and J. A. Herrera-Cervera for technical assistance. We are very grateful to all the scientists who sent us strains, and also to the referees who significantly contributed to improving this manuscript.

This work was funded by INRA as part of AIP Microbiologie, and by MESR as part of ACC-SV no. 6. It was also financially supported by a grant from the Conseil Régional de Bourgogne.

REFERENCES

Amarger, N., Bours, M., Revoy, F., Allard, M. R. & Laguerre, G. (1994). *Rhizobium tropici* nodulates field-grown *Phaseolus vulgaris* in France. *Plant Soil* **161**, 147–156.

Amarger, N., Macheret, V. & Laguerre, G. (1997). *Rhizobium gallicum* sp. nov. and *Rhizobium giardinii* sp. nov. from *Phaseolus vulgaris* nodules. *Int J Syst Bacteriol* **47**, 996–1006.

Bergersen, F. J., Turner, G. L., Amarger, N., Mariotti, N. & Mariotti, A. (1986). Strain of *Rhizobium lupini* determines natural abundance of 15N in root nodules of *Lupinus* spp. *Soil Biol Biochem* **18**, 97–101.

van Berkum, P., Beyene, D. & Eardly, B. D. (1996). Phylogenetic relationships among *Rhizobium* species nodulating the common bean (*Phaseolus vulgaris* L.). *Int J Syst Bacteriol* **46**, 240–244.

van Berkum, P., Beyene, D., Bao, G., Campbell, T. A. & Eardly, B. D. (1998). *Rhizobium mongolense* sp. nov. is one of three rhizobial genotypes identified which nodulate and form nitrogen-fixing symbioses with *Medicago ruthenica* [(L.) Ledebour]. *Int J Syst Bacteriol* **48**, 13–22.

Chen, W., Wang, E., Wang, S., Li, Y., Chen, X. & Li, Y. (1995). Characteristics of *Rhizobium tianshanense* sp. nov., a moderately and slowly growing root nodule bacterium isolated from an arid saline environment in Xianjiang, People's Republic of China. *Int J Syst Bacteriol* **45**, 153–159.

Del Papa, M. F., Balagué, L. J., Castro Sowinski, S. & 9 other authors (1999). Isolation and characterization of alfalfa-nodulating rhizobia present in acidic soils of central Argentina and Uruguay. *Appl Environ Microbiol* **65**, 1420–1427.

Dessen, P., Fondrat, C., Valencien, C. & Mugnier, C. (1990). Bisance: a French service for access to biomolecular sequence databases. *Comput Appl Biosci* **6**, 355–356.

Dobert, R. C., Breil, B. T. & Triplett, E. W. (1994). DNA sequence of the common nodulation genes of *Bradyrhizobium elkanii* and their phylogenetic relationships to those of other nodulating bacteria. *Mol Plant–Microbe Interact* **7**, 564–572.

Eardly, B. D., Young, J. P. W. & Selander, R. K. (1992). Phylogenetic position of *Rhizobium* sp. strain Or 191, a symbiont of both *Medicago sativa* and *Phaseolus vulgaris*, based on partial sequences of the 16S rRNA and *nif*H genes. *Appl Environ Microbiol* **58**, 1809–1815.

Felsenstein, J. (1989). PHYLIP – phylogeny inference package (version 3.2). *Cladistics* **5**, 164–166.

Garcia-de los Santos, A., Brom, S. & Romero, D. (1996). *Rhizobium* plasmids in bacteria-legume interactions. *World J Microbiol Biotechnol* **12**, 119–125.

Geniaux, E., Laguerre, G. & Amarger, N. (1993). Comparison of geographically distant populations of *Rhizobium* isolated from root nodules of *Phaseolus vulgaris*. *Mol Ecol* **2**, 295–302.

Graham, P. H. (1976). Identification and classification of root nodule bacteria. In *Symbiotic Nitrogen Fixation in Plants*, pp. 99–112. Edited by P. S. Nutman. International Biological Programme No. 7. London: Cambridge University Press.

Graham, P. H., Sadowsky, M. J., Keyser, H. H. & 8 other authors (1991). Proposed minimal standards for the description of new genera and species of root- and stem-nodulating bacteria. *Int J Syst Bacteriol* **41**, 582–587.

Haukka, K., Lindström, K. & Young, J. P. W. (1998). Three phylogenetic groups of *nodA* and *nifH* genes in *Sinorhizobium* and *Mesorhizobium* isolated from leguminous trees growing in Africa and Latin America. *Appl Environ Microbiol* **64**, 419–426.

Hennecke, H., Kaluza, K., Thöny, B., Fuhrmann, M., Ludwig, W. & Stackebrandt, E. (1985). Concurrent evolution of nitrogenase genes and 16S rRNA in *Rhizobium* species and other nitrogen fixing bacteria. *Arch Microbiol* **142**, 342–348.

Herrera-Cervera, J. A., Caballero-Medallo, J., Laguerre, G., Tichy, H. V., Requena, N., Amarger, N., Martinez-Romero, E., Olivares, J. & Sanjuan, J. (1999). At least five rhizobial species nodulate *Phaseolus vulgaris* in a Spanish soil. *FEMS Microbiol Ecol* **30**, 87–97.

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* **16**, 111–120.

Kuykendall, L. D., Saxena, B., Devine, T. E. & Udell, S. E. (1992). Genetic diversity in *Bradyrhizobium japonicum* Jordan 1982 and a proposal for *Bradyrhizobium elkanii* sp. nov. *Can J Microbiol* **38**, 501–505.

Lagacherie, B., Hugot, R. & Amarger, N. (1977). Sélection de souches de *Rhizobium japonicum* d'après leur compétitivité pour l'infection. *Ann Agron* **28**, 379–389.

Lagacherie, B., Bours, M., Giraud, J. J. & Sommer, G. (1983). Interactions entre différentes souches de *Rhizobium lupini* et les espèces ou cultivars de lupin (*Lupinus albus*, *Lupinus luteus* et *Lupinus mutabilis*). *Agronomie* **3**, 809–816.

Laguerre, G., Allard, M. R., Revoy, F. & Amarger, N. (1994). Rapid identification of rhizobia by restriction fragment length polymorphism analysis of PCR-amplified 16S rRNA genes. *Appl Environ Microbiol* **60**, 56–63.

Laguerre, G., Mavingui, P., Allard, M. R., Charnay, M. P., Louvrier, P., Mazurier, S. I., Rigottier-Gois, L. & Amarger, N. (1996). Typing of rhizobia by PCR DNA fingerprinting and PCR-restriction fragment length polymorphism analysis of chromosomal and symbiotic gene regions: application to *Rhizobium leguminosarum* and its different biovars. *Appl Environ Microbiol* **62**, 2029–2036.

Laguerre, G., van Berkum, P., Amarger, N. & Prévost, D. (1997). Genetic diversity of rhizobial symbionts isolated from legume species within the genera *Astragalus*, *Oxytropis* and *Onobrychis*. *Appl Environ Microbiol* **63**, 4748–4758.

de Lajudie, P., Willems, A., Pot, B. & 7 other authors (1994). Polyphasic taxonomy of rhizobia: emendation of the genus

Sinorhizobium, and description of *Sinorhizobium meliloti* comb. nov., *Sinorhizobium saheli* sp. nov., and *Sinorhizobium teranga* sp. nov. *Int J Syst Bacteriol* **44**, 715–733.

de Lajudie, P., Laurent-Fulele, E., Willems, A., Torck, U., Coopman, R., Collins, M. D., Kersters, K., Dreyfus, B. & Gillis, M. (1998a). *Allorhizobium undicola* gen. nov., sp. nov., nitrogen-fixing bacteria that efficiently nodulate *Neptunia natans* in Senegal. *Int J Syst Bacteriol* **48**, 1277–1290.

de Lajudie, P., Willems, A., Nick, G. & 9 other authors (1998b). Characterization of tropical tree rhizobia and description of *Mesorhizobium plurifarium* sp. nov. *Int J Syst Bacteriol* **48**, 369–382.

Lindström, K. (1989). *Rhizobium galegae*, a new species of legume root nodule bacteria. *Int J Syst Bacteriol* **39**, 365–367.

Lindström, K., Paulin, L., Roos, C. & Suominen, L. (1995). Nodulation genes of *Rhizobium galegae*. In *Nitrogen Fixation: Fundamentals and Applications. Proceedings of the 10th International Congress on Nitrogen Fixation*, pp. 365–370. Edited by I. A. Tikhonovitch, N. A. Provorov, V. I. Romanov & W. E. Nexton. Dordrecht: Kluwer.

Lortet, G., Méar, N., Lorquin, J., Dreyfus, B., de Lajudie, P., Rosenberg, C. & Boivin, C. (1996). Nod factor thin-layer chromatography profiling as a tool to characterize symbiotic specificity of rhizobial strains: application to *Sinorhizobium saheli*, *S. teranga*, and *Rhizobium* sp. strains isolated from *Acacia* and *Sesbania*. *Mol Plant–Microbe Interact* **9**, 736–747.

Martinez, E., Pardo, M. A., Palacios, R. & Cevallos, M. A. (1985). Reiteration of nitrogen fixation gene sequences and specificity to *Rhizobium* in nodulation and nitrogen fixation in *Phaseolus vulgaris*. *J Gen Microbiol* **131**, 1779–1786.

Martinez-Romero, E. & Caballero-Mellado, J. (1996). *Rhizobium* phylogenies and bacterial genetic diversity. *Crit Rev Plant Sci* **15**, 113–140.

Martinez-Romero, E., Segovia, L., Mercante, F. M., Franco, A. A., Graham, P. & Pardo, M. A. (1991). *Rhizobium tropici*, a novel species nodulating *Phaseolus vulgaris* L. beans and *Leucaena* sp. trees. *Int J Syst Bacteriol* **41**, 417–426.

Michiels, J., Dombrecht, B., Vermeiren, N., Xi, C., Luyten, E. & Venderleyden, J. (1998). *Phaseolus vulgaris* is a non-selective host for nodulation. *FEMS Microbiol Ecol* **26**, 193–205.

Moulin, L., Debellé, F., Giraud, E., Mangin, B., Dénarié, J. & Boivin-Masson, C. (2000). The *nodA* sequence of rhizobia gives clues on structural features of Nod factors. In *Nitrogen Fixation: From Molecules to Crop Productivity*, p. 204. Edited by F. O. Pedrosa, M. Hungria, M. G. Yates & W. E. Newton. Dordrecht: Kluwer.

Perret, X., Staehelin, C. & Broughton, W. J. (2000). Molecular basis of symbiotic promiscuity. *Microbiol Mol Biol Rev* **64**, 180–201.

Poupot, R., Martinez-Romero, E. & Promé, J. C. (1993). Nodulation factors from *Rhizobium tropici* are sulfated or non-sulfated chitopentasaccharides containing an *N*-methyl-*N*-acyl-glucosamine terminus. *Biochemistry* **32**, 10430–10435.

Poupot, R., Martinez-Romero, E., Gautier, N. & Promé, J. C. (1995). Wild type *Rhizobium etli*, a bean symbiont, produces acetyl-fucosylated, N-methylated, and carbamoylated nodulation factors. *J Biol Chem* **270**, 6050–6055.

Prévost, D., Macheret, V. & Laguerre, G. (2000). Phylogenetic comparisons of symbiotic (*nodC* and *nifH*) and 16S rDNA genes in strains of *Rhizobium*, *Mesorhizobium* and *Bradyrhizobium* isolated from *Astragalus*, *Oxytropis* and *Onobrychis* spp. In *Nitrogen Fixation: From Molecules to Crop Productivity*, p. 205. Edited by F. O. Pedrosa, M. Hungria, M. G. Yates & W. E. Newton. Dordrecht: Kluwer.

Pueppke, S. G. & Broughton, W. J. (1999). *Rhizobium* sp. strain NGR234 and *R. fredii* USDA257 share exceptionally broad, nested host ranges. *Mol Plant–Microbe Interact* **12**, 293–318.

Romero, D., Davila, G. & Palacios, R. (1998). The dynamic genome of Rhizobium. In *Bacterial Genomes: Physical Structure and Analysis*, pp. 153–161. Edited by F. J. de Bruijn, J. R. Lupski & G. M. Weinstock. New York: Chapman & Hall, International Thompson Publishing.

Segovia, L., Young, J. P. W. & Martinez-Romero, E. (1993). Reclassification of American *Rhizobium leguminosarum* biovar phaseoli type I strains as *Rhizobium etli* sp. nov. *Int J Syst Bacteriol* **43**, 374–377.

Sessitsch, A., Ramirez-Saad, H., Hardarson, G., Akkermans, A. D. I. & de Vos, W. M. (1997). Classification of Austrian rhizobia and the Mexican isolate FL27 obtained from *Phasolus vulgaris* L. as *Rhizobium gallicum*. *Int J Syst Bacteriol* **47**, 1097–1101.

Sullivan, J. T. & Ronson, C. W. (1998). Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into a phe-tRNA gene. *Proc Natl Acad Sci U S A* **95**, 5145–5149.

Sullivan, J. T., Patrick, H. N., Lowther, W. L., Scott, D. B. & Ronson, C. W. (1995). Nodulating strains of *Rhizobium loti* arise through chromosomal symbiotic gene transfer in the environment. *Proc Natl Acad Sci U S A* **92**, 8985–8989.

Sullivan, J. T., Eardly, B. D., van Berkum, P. & Ronson, C. W. (1996). Four unnamed species of nonsymbiotic rhizobia isolated from the rhizosphere of *Lotus corniculatus*. *Appl Environ Microbiol* **62**, 2818–2825.

Thompson, J. D., Higgins, C. W. & Gibson, T. J. (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.

Turner, S. L. & Young, J. P. W. (2000). The glutamine synthetases of rhizobia: phylogenetics and evolutionary implications. *Mol Biol Evol* **17**, 309–319.

Ueda, T., Suga, Y., Yahiro, N. & Matsuguchi, T. (1995). Phylogeny of Sym plasmids of rhizobia by PCR-based sequencing of a *nodC* segment. *J Bacteriol* **177**, 468–472.

Wang, E. T., van Berkum, P., Beyene, D., Sui, X. H., Dorado, O., Chen, W. X. & Martinez-Romero, E. (1998). *Rhizobium huautlense* sp. nov., a symbiont of *Sesbania herbacea* that has a close phylogenetic relationship with *Rhizobium galegae*. *Int J Syst Bacteriol* **48**, 687–699.

Wang, E. T., Rogel, A., Garcia-de los Santos, A., Martinez-Romero, J., Cevallos, M. A. & Martinez-Romero, E. (1999a). *Rhizobium etli* bv. mimosae, a novel biovar isolated from *Mimosa affinis*. *Int J Syst Bacteriol* **49**, 1479–1491.

Wang, E. T., van Berkum, P., Sui, X. H., Beyene, D., Chen, W. X. & Martinez-Romero, E. (1999b). Diversity of rhizobia associated with *Amorpha fruticosa* isolated from Chinese soils and description of *Mesorhizobium amorphae* sp. nov. *Int J Syst Bacteriol* **49**, 51–65.

Wernegreen, J. J. & Riley, M. A. (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.

Xu, L. M., Ge, C., Cui, Z. & Fan, H. (1995). *Bradyrhizobium liaoningense* sp. nov. isolated from the root nodules of soybeans. *Int J Syst Bacteriol* **45**, 706–711.

Yang, G. P., Debellé, F., Savagnac, A. & 9 other authors (1999). Structure of the *Mesorhizobium huakuii* and *Rhizobium galegae*

Nod factors: a cluster of phylogenetically related legumes are nodulated by rhizobia producing Nod factors with α,β-unsaturated *N*-acyl substitutions. *Mol Microbiol* **31**, 227–237.

Young, J. P. W. (1985). *Rhizobium* population genetics: enzyme polymorphism in isolates from peas, clover, beans and lucerne grown at the same site. *J Gen Microbiol* **131**, 2399–2408.

Young, J. P. W. (1992). Phylogenetic classification of nitrogen-fixing organisms. In *Biological Nitrogen Fixation*, pp. 43–86. Edited by G. Stacey, R. H. Burris & H. J. Evans. New York: Chapman & Hall.

Young, J. P. W. & Haukka, K. E. (1996). Diversity and phylogeny of rhizobia. *New Phytol* **133**, 87–94.

Zhang, X. X., Turner, S. L., Guo, X. W., Yang, H. J., Debellé, F., Yang, G. P., Dénarié, J., Young, J. P. W. & Li, F. D. (2000). The common nodulation genes of *Astragalus sinicus* rhizobia are conserved despite chromosomal diversity. *Appl Environ Microbiol* **66**, 2988–2995.

Received 12 July 2000; revised 28 November 2000; accepted 15 December 2000.