Phylogeny of nodulation and nitrogen-fixation genes in *Bradyrhizobium*: supporting evidence for the theory of monophyletic origin, and spread and maintenance by both horizontal and vertical transfer

Pâmela Menna[1,2] and Mariangela Hungria[1,2]

[1]Embrapa Soja, Cx Postal 231, 86001-970 Londrina, Paraná, Brazil

[2]Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq-MCT), Brasilia, Federal District, Brazil

Correspondence
Mariangela Hungria
mariangela.hungria@embrapa.br
or hungria@pq.cnpq.br

Bacteria belonging to the genus *Bradyrhizobium* are capable of establishing symbiotic relationships with a broad range of plants belonging to the three subfamilies of the family Leguminosae (=Fabaceae), with the formation of specialized structures on the roots called nodules, where fixation of atmospheric nitrogen takes place. Symbiosis is under the control of finely tuned expression of common and host-specific nodulation genes and also of genes related to the assembly and activity of the nitrogenase, which, in *Bradyrhizobium* strains investigated so far, are clustered in a symbiotic island. Information about the diversity of these genes is essential to improve our current poor understanding of their origin, spread and maintenance and, in this study, we provide information on 40 *Bradyrhizobium* strains, mostly of tropical origin. For the nodulation trait, common (*nodA*), *Bradyrhizobium*-specific (*nodY/K*) and host-specific (*nodZ*) nodulation genes were studied, whereas for fixation ability, the diversity of *nifH* was investigated. In general, clustering of strains in all *nod* and *nifH* trees was similar and the *Bradyrhizobium* group could be clearly separated from other rhizobial genera. However, the congruence of *nod* and *nif* genes with ribosomal and housekeeping genes was low. *nodA* and *nodY/K* were not detected in three strains by amplification or hybridization with probes using *Bradyrhizobium japonicum* and *Bradyrhizobium elkanii* type strains, indicating the high diversity of these genes or that strains other than photosynthetic *Bradyrhizobium* must have alternative mechanisms to initiate the process of nodulation. For a large group of strains, the high diversity of *nod* genes (with an emphasis on *nodZ*), the low relationship between *nod* genes and the host legume, and some evidence of horizontal gene transfer might indicate strategies to increase host range. On the other hand, in a group of five symbionts of *Acacia mearnsii*, the high congruence between *nod* and ribosomal/housekeeping genes, in addition to shorter *nodY/K* sequences and the absence of *nodZ*, highlights a co-evolution process. Additionally, in a group of *B. japonicum* strains that were symbionts of soybean, vertical transfer seemed to represent the main genetic event. In conclusion, clustering of *nodA* and *nifH* gives additional support to the theory of monophyletic origin of the symbiotic genes in *Bradyrhizobium* and, in addition to the analysis of *nodY/K* and *nodZ*, indicates spread and maintenance of *nod* and *nif* genes through both vertical and horizontal transmission, apparently with the dominance of one or other of these events in some groups of strains.

INTRODUCTION

Some prokaryotes have a remarkable capacity to fix atmospheric nitrogen, thus providing it in a utilizable form (ammonia) to plants. The capacity to fix nitrogen is determined by a highly conserved enzyme complex called nitrogenase, which is inactivated in the presence of oxygen (Zehr *et al.*, 1993; Angus & Hirsch, 2010). Nitrogenase probably arose in the Archean age and, throughout evolution, has been maintained in several genera that are collectively known as diazotrophic micro-organisms (Dixon & Kahn, 2004). Diazotrophs are found in a variety of phylogenetic groups such as green sulphur bacteria, firmibacteria, actinomycetes, cyanobacteria and all subdivisions of

Abbreviation: LCOs, lipo-chitin oligosaccharides.

The GenBank/EMBL/DDBJ accession numbers for the sequences reported in this paper are given in Figs 1–4.

Two supplementary figures and a table are available with the online version of this paper.

the *Proteobacteria*, and, in *Archaea*, the nitrogen fixation trait predominates in the methanogenic group (Dixon & Kahn, 2004; Lloret & Martínez-Romero, 2005; Angus & Hirsch, 2010).

The symbiotic association between diazotrophic bacteria commonly known as rhizobia with plants probably arose later, with the emergence of the first terrestrial plants, and was established mainly with members of the family Leguminosae (=Fabaceae), which arose 60–65 million years ago. Symbiotic nitrogen fixation is a highly complex process and, in bacteria, involves the expression of several genes, including those related to the assembly and activity of the nitrogenase (*nif* and *fix*) and others (*nod*, *nol* and *noe*) related to the formation of highly specialized structures, the nodules, where this process takes place. These genes are clustered in complex operons located either in plasmids (e.g. in *Rhizobium* species, *Sinorhizobium* species, *Mesorhizobium amorphae*, *Mesorhizobium huakuii*) or in symbiotic islands flanked by insertion sequences (e.g. in *Mesorhizobium loti* and *Bradyrhizobium japonicum*) (Freiberg *et al.*, 1997; Göttfert *et al.*, 2001; Kaneko *et al.*, 2002; Lloret & Martínez-Romero, 2005).

Nodulation genes are responsible for the production of chitooligosaccharides, also called Nod factors. Usually Nod factors are responsible for the specificity between plants and bacteria and consist of chitin oligomers mono *N*-acylated at the non-reducing end with distinct substituents at both ends of the molecule; in the majority of rhizobia studied so far, synthesis of these molecules is controlled by the *nodABC* genes. In addition, distinct rhizobia carry different combinations of nodulation genes that were probably recruited from paralogues during the course of evolution and are named 'host-specific nodulation genes' (*hsn*) (Lerouge *et al.*, 1990; Broughton *et al.*, 2000; Lloret & Martínez-Romero, 2005). It is generally believed that natural events of lateral gene transfer and duplication of nodulation genes have contributed to the evolution and spread of symbiotic ability throughout several genera of bacteria (Barcellos *et al.*, 2007; Zhao *et al.*, 2008).

The genus *Bradyrhizobium* encompasses diazotrophic bacteria that can live in symbiotic and endophytic association with legumes and non-legumes, and are characterized by physiological and symbiotic versatility and broad geographic distribution. *Bradyrhizobium* is the most abundant rhizobial group identified in root nodules of legumes growing in tropical and subtropical areas and the genus is also well-known for its broad host range, which is an adaptation allowing persistence in tropical areas that are known for their high legume diversity. In previous studies by our group, symbionts from plants belonging to the three subfamilies of the family Leguminosae have demonstrated great genetic diversity in their 16S rRNA and other housekeeping genes (Menna *et al.*, 2006, 2009a, b). It has also been shown that analysis of nitrogen fixation and nodulation genes may provide valuable information about the evolution of *Bradyrhizobium* (Stępkowski *et al.*, 2005; Stępkowski *et al.*, 2007; Steenkamp *et al.*, 2008). For example, studies of *nodA*, *nodZ*, *nolL* and *noeI* genes in *Bradyrhizobium* have shown a monophyletic origin and comparison with housekeeping genes has indicated that the spread and maintenance of the *nod* genes seem to have occurred mainly through vertical transmission, although lateral gene transfer might also have played a significant role (Moulin *et al.*, 2004; Stępkowski *et al.*, 2005, 2007; Steenkamp *et al.*, 2008).

In this study, the genetic diversity of nodulation and nitrogen-fixation genes was analysed in 40 *Bradyrhizobium* strains, symbionts isolated from a variety of legume species, mostly of tropical origin. For the nodulation trait, common and host-specific genes (*nodY/K*, *nodA*, *nodZ*) were examined whereas, for nitrogen-fixation ability, the diversity of *nifH* was investigated. In addition to showing high diversity of *nod* and *nif* genes, our results reveal a complex evolutionary pattern for tropical *Bradyrhizobium* species.

METHODS

Strains. Forty *Bradyrhizobium* SEMIA strains from the Brazilian *Rhizobium* Culture Collection SEMIA of the FEPAGRO-MIRCEN [Fundação Estadual de Pesquisa Agropecuária (Rio Grande do Sul, Brazil) – Microbiological Resources Center] (IBP World Catalogue of *Rhizobium* Collections no. 443 in the World Federation of Culture Collections) were used in this study (Table 1). The strains were isolated from members of the three subfamilies and ten tribes of the family Leguminosae (=Fabaceae) (Table 1). The strains are deposited in the Diazotrophic and Plant Growth Promoting Bacteria Culture Collection of Embrapa Soja (http://www.bmrc.lncc.br). Preparation of stock cultures, strain growth conditions and maintenance were as described by Menna *et al.* (2006).

DNA extraction, amplification and sequencing of *nod* and *nif* genes. Total genomic DNA of each strain was extracted from bacterial batch cultures grown in YM broth until late exponential phase (10^9 cells ml^{-1}) and extraction of DNA was performed as described previously (Menna *et al.*, 2006). DNA from each strain was amplified by using specific primers for the regions coding for *nodY/K*, *nodA*, *nodZ* and *nifH* genes (Table 2). For each PCR, 40 ng DNA was used; primers, amplification conditions and references are listed in Table 2. PCR products were purified using the QIAquick PCR Purification kit (Qiagen) and sequencing was performed as described by Menna *et al.* (2006). Some PCR products (*nifH* of SEMIA 6156, SEMIA 6434 and SEMIA 6014) had to be cloned with the pGEM Easy Vector Systems (Promega), according to the manufacturer's instructions, before sequencing.

Cluster analyses. After sequencing, the nucleotide sequences generated were analysed with the programs Phred, Phrap and Consed, as described previously (Menna *et al.*, 2009a). The sequences obtained were confirmed in the 5′ and 3′ directions and submitted to GenBank; accession numbers are listed in Figs 1–4. Sequences were analysed using the software MEGA version 4.0 with default parameters, the K2P distance model (Kimura, 1980) and the neighbour-joining algorithm (Saitou & Nei, 1987). Statistical support for the tree nodes was evaluated by bootstrap analyses (Felsenstein, 1985) with 1000 samplings (Hedges, 1992). Accession numbers of reference/type strains used for alignment and comparison are also listed in Figs 1–4.

Table 1. *Bradyrhizobium* strains used in this study

Strain	Classification-based MLSA*	Biovar	Other designations	Host plant†	Common name	Geographic origin
SEMIA 511	*B. japonicum*	*glycinearum*	UW 511, USDA 500	*Glycine max* (L.) Merr.[3,7]	Soybean	USA
SEMIA 512	*B. japonicum*	*glycinearum*	3I1b73 (USDA-Dr Erdman)	*Glycine max* (L.) Merr.[3,7]	Soybean	USA
SEMIA 560	*Bradyrhizobium* sp.	*glycinearum*	No synonyms	*Glycine max* (L.) Merr.[3,7]	Soybean	Brazil
SEMIA 587	*B. elkanii*		BR 96	*Glycine max* (L.) Merr.[3,7]	Soybean	Brazil
SEMIA 656	*Bradyrhizobium* sp.		Original SEMIA	*Neonotonia wightii* (Wight & Arn.) J. A. Lackey[3,7]	Perennial soybean	Brazil
SEMIA 695	*B. elkanii*		E 85, QA 922, SU 422, NA 630	*Neonotonia wightii* (Wight & Arn.) J. A. Lackey[3,7]	Perennial soybean	Australia
SEMIA 928	*B. canariense*	*genistearum*	W-72	*Lupinus* sp.[3,13]	Lupin	Not known
SEMIA 5011	*B. elkanii*		Original SEMIA	*Glycine max* (L.) Merr.[3,7]	Soybean	Brazil
SEMIA 5019	*B. elkanii*		29 W, BR 29	*Glycine max* (L.) Merr.[3,7]	Soybean	Brazil
SEMIA 5025	*B. liaoningense*	*glycinearum*	TAL 411, Tha3	*Glycine max* (L.) Merr.[3,7]	Soybean	Thailand
SEMIA 5026	*B. elkanii*		TAL 415, THA 9	*Glycine max* (L.) Merr.[3,7]	Soybean	Thailand
SEMIA 5027	*B. elkanii*		TAL 183, 61a76	*Glycine max* (L.) Merr.[3,7]	Soybean	USA
SEMIA 5045	*B. japonicum*	*glycinearum*	Nit 123P35	*Glycine max* (L.) Merr.[3,7]	Soybean	Not known
SEMIA 5062	*B. liaoningense*		SVJ-04	*Glycine max* (L.) Merr.[3,7]	Soybean	Brazil
SEMIA 5079	*B. japonicum*	*glycinearum*	CPAC 15, DF 24	*Glycine max* (L.) Merr.[3,7]	Soybean	Brazil
SEMIA 5080	*B. japonicum*	*glycinearum*	CPAC 7	*Glycine max* (L.) Merr.[3,7]	Soybean	Brazil
SEMIA 6014	*Bradyrhizobium* sp.		CIAT 1316, USM 128, TAL 405	*Stylosanthes guianensis* (Aubl.) Sw.[3,5]	Stylosanthes	Peru
SEMIA 6028	*Bradyrhizobium* sp.		TAL 569, SPRL 472, MAR 472	*Desmodium uncinatum* (Jacq.) DC.[3,11]	Silverleaf Desmodium	Zimbabwe
SEMIA 6053	*Bradyrhizobium* sp.		TAL 827, UMKI 28	*Clitoria ternatea* L.[3,7]	Blue-pea	Malaysia
SEMIA 6059	*B. japonicum*		USDA 3309	*Psophocarpus tetragonolobus* (L.) DC.[3,7]	Winged bean	USA
SEMIA 6069	*Bradyrhizobium* sp.		DF 10	*Leucaena leucocephala* (Lam.)[1,12]	Jumbie bean	Brazil
SEMIA 6077	*B. yuanmingense*		CB 82	*Stylosanthes* sp.[3,5]	Stylosanthes	Australia
SEMIA 6093	*Bradyrhizobium* sp.		USDA 3331	*Aeschynomene americana* L.[3,5]	Shyleaf	USA
SEMIA 6099	*Bradyrhizobium* sp.		BR 5004, LMG 9989	*Dimorphandra exaltata* Schott[2,10]	Dimorphandra	Brazil
SEMIA 6101	*Bradyrhizobium* sp.		BR 8404	*Dalbergia nigra* (Vell.) Benth.[3,9]	Brazilian rosewood	Brazil
SEMIA 6146	*Bradyrhizobium* sp.		BR 1808	*Centrosema* sp.[3,7]	Centrosema	Brazil
SEMIA 6148	*Bradyrhizobium* sp.		SMS 303	*Neonotonia wightii* (Wight & Arn.) J. A. Lackey[3,7]	Perennial soybean	Brazil
SEMIA 6152	*Bradyrhizobium* sp.		BR-1602	*Calopogonium* sp.[3,7]	Calopogonium	Brazil
SEMIA 6156	*Bradyrhizobium* sp.		CPAC IJ	*Crotalaria spectabilis* Roth[3,8]	Rattlebox	Brazil
SEMIA 6160	*Bradyrhizobium* sp.		BR 5610	*Albizia lebbeck* (L.) Benth.[1,6]	Rain tree	Brazil
SEMIA 6163	*Bradyrhizobium* sp.		BR 3607	*Acacia mearnsii* De Wild.[1,4]	Black Wattle	Brazil
SEMIA 6164	*Bradyrhizobium* sp.		BR 3608, LMG 9960	*Acacia mearnsii* De Wild.[1,4]	Black Wattle	Brazil
SEMIA 6179	*Bradyrhizobium* sp.		Original SEMIA	*Acacia mearnsii* De Wild.[1,4]	Black Wattle	Brazil
SEMIA 6186	*Bradyrhizobium* sp.		Original SEMIA	*Acacia mearnsii* De Wild.[1,4]	Black Wattle	Brazil
SEMIA 6187	*Bradyrhizobium* sp.		Original SEMIA	*Acacia mearnsii* De Wild.[1,4]	Black Wattle	Brazil
SEMIA 6192	*Bradyrhizobium* sp.		Original SEMIA	*Tipuana tipu*[3,9]	Tipu	Brazil
SEMIA 6319	*B. yuanmingense*		NC 92	*Arachis* sp.[3,5]	Arachis	Bolivia

Table 1. cont.

Strain	Classification-based MLSA*	Biovar	Other designations	Host plant†	Common name	Geographic origin
SEMIA 6374	*Bradyrhizobium* sp.		Not known	*Arachis pintoi* Krapov. & W. C. Greg[3,5]	Forage peanut	Not known
SEMIA 6434	*Bradyrhizobium* sp.		Not known	*Inga* sp.[1,6]	Inga	Not known
SEMIA 6440	*Bradyrhizobium* sp.		MGAP 13	*Arachis pintoi* Krapov. & W. C. Greg[3,5]	Forage peanut	Brazil

*According to Menna *et al.* (2009b).
†Classifications designated as follows: [1], subfamily Mimosoideae; [2], subfamily Caesalpinioideae; [3], subfamily Papilionoideae; [4], tribe Acacieae; [5], tribe Aeschynomeneae; [6], tribe Ingeae; [7], tribe Phaseoleae; [8], tribe Crotalarieae; [9], tribe Dalbergieae; [10], tribe Caesalpinieae; [11], tribe Desmodieae; [12], tribe Mimoseae; and [13], tribe Genisteae.

Southern hybridization assays. Each DNA sample (500 ng) was digested overnight with 2 µl *Eco*RI restriction endonuclease (10 U μl^{-1}; Invitrogen), separated in 1 % agarose gel and then transferred to a Hybond-N+ membrane (Amersham Biosciences) by the Southern blot standard procedure (Sambrook & Russell, 2001). Labelling of probes and hybridization procedures were performed using the AlkPhos Direct labelling reagents (Amersham Biosciences), according to the manufacturer's instructions. The *nodZ*, *nodA* and *nodY/K* partial genes from *Bradyrhizobium japonicum* strain USDA 6^T and *Bradyrhizobium elkanii* strain USDA 76^T, amplified as described in Table 2, were used as probes. *Escherichia coli* TOP10 was used as control for the hybridization assay. After hybridization, detection was carried out using CDP-*Star* reagent in a Hypercassette (both Amersham Biosciences), according to the manufacturer's instructions.

RESULTS

Ribosomal and housekeeping genes

To allow a comparison with nodulation and nitrogen fixation genes, phylogenetic trees built with 16S rRNA genes (Supplementary Fig. S1, available in IJSEM Online) and with the concatenated housekeeping genes *dnaK*, *glnII* and *recA* (Supplementary Fig. S2, available in IJSEM Online) were included. The trees were rebuilt using the 40 strains from this study and sequences previously obtained by Menna *et al.* (2009a). In the 16S rRNA gene phylogenetic tree, strains were split into two well supported clusters (I and II), with the majority of the strains (23 out of 40) being related to *B. japonicum*, *Bradyrhizobium yuanmingense*, *Bradyrhizobium liaoningense*, *Bradyrhizobium betae* and *Bradyrhizobium canariense* type/reference strains. Strains in cluster II showed high similarity to the type strains of *Bradyrhizobium jicamae*, *Bradyrhizobium pachyrhizi* and *B. elkanii* (Supplementary Fig. S1). In both clusters I and II of the 16S rRNA gene tree, subdivision into subclusters was unclear; therefore, phylogeny of housekeeping genes was also performed. The tree built with the three housekeeping genes also split the strains into two main clusters (I and II); however, subclusters were more clearly observed (Table 3; Supplementary Fig. S2).

Nodulation genes

No amplification products were formed with the TSnodD1-1a/TSnodB1 primers for three of the 40 strains: SEMIA 6014 (from *Stylosanthes guianensis*), SEMIA 6192 (from *Tipuana tipu*) and SEMIA 6434 (from *Inga* sp.). For the remaining strains, fragments of about 2 kb containing the whole *nodD–nodA* intergenic region (i.e. *nod* box and *nodY/K*), the *nodA* gene and 230–530 bp of the *nodB* gene were obtained. Shorter fragments of about 1800 bp were obtained for strains SEMIA 6186, 6164, 6163, 6187 and 6179, all symbionts of *Acacia mearnsii*. The 2 kb fragments obtained from PCR amplification were submitted to a new PCR cycle with the primer set TSnodD1-1a/TSnodA2, resulting in fragments of about 1200 bp, except again for SEMIA strains 6186, 6164, 6163, 6187 and 6179, which resulted in fragments of about 800 bp but, in both cases,

Table 2. Primers and DNA amplification conditions used in this study

Primer	Sequence (5′→3′)*	Target gene (position)†	Amplified genes in PCR	PCR cycling	Reference
TSnodD1-1a TSnodB1	CAGATCNAGDCCBTTGAARCGCA AGGATAYCCGTCG TGCAGGAGCA	*nodD1* (24–2) *nodB* (534–512)	*nod* box§, *nodY/K*§, *nodA*§	2 min 94 °C, 20 × (1 min 94 °C, 2 min using temperature rescinding from 60 to 50 °C, 2 min 72 °C), 7 min 72 °C	Moulin *et al.* (2004)
TSnodD1-1a‡ TSnodA2‡	CAGATCNAGDCCBTTGAARCGCA GCTGATTCCAAGBCCYTCVAGATC	*nodD1* (24–2) *nodA* (348–325)	*nod* box§, *nodY/K*§, *nodA*‖	1 min 95 °C, 29 × (45 s 95 °C, 30 s 58 °C, 1.5 min 72 °C), 5 min 72 °C	This study (second PCR)
TSnodZ3‡ TSnodZ4‡	GGTTTCGGYGAYTGYCTBTGGTC AATRTCTTCGCCRTTRCCRTGCC	*nodZ* (40–62) *nodZ* (552–530)	*nodZ*‖	2 min 95 °C, 35 × (45 s 94 °C, 30 s 53 °C, 1.5 min 72 °C), 7 min 72 °C	Moulin *et al.* (2004)
nifHF‡ nifHI‡	TACGGNAARGGSGGNATCGGCAA AGCATGTCYTCSAGYTCNTCCA	*nifH* (25–50) *nifH* (787–808)	*nifH*‖	2 min 95 °C, 35 × (1 min 94 °C, 45 s 60 °C, 2 min 72 °C), 7 min 72 °C	Laguerre *et al.* (2001)

*Mixtures of bases used at certain positions are given as: S, G or C; Y, C or T; R, A or G; N, A, T, C or G.
†For *nod* genes, primer positions correspond to sequences in *B. japonicum* USDA 110; for *nifH*, primer positions correspond to sequences in *R. leguminosarum* bv. *trifolii*.
‡Primers also used for sequencing reaction.
§Complete sequence.
‖Partial sequence.

the fragments contained *nodY/K* and about 300 bp of *nodA*. Finally, the *nodY/K* sequences of these five strains were approximately 300 bp shorter and therefore only a 402–514 bp consensus fragment present in all strains was considered in the phylogenetic analysis.

The phylogenetic tree built with the partial sequences of *nodA* (272–277 bp) split the strains into six well-supported major clusters (I–VI) (Fig. 1; Table 3). Cluster I grouped together 25 SEMIA strains and the reference/type strains used in this study (Fig. 1) with bootstrap support of 66 %. Six subclusters (I.1 to I.6) were formed in cluster I. Subcluster I.1 included seven SEMIA strains of distinct geographic origin, all symbionts of *Glycine max*, and also the type/reference strains of *B. japonicum* USDA 6^T and USDA 110. Subcluster I.2 included strain SEMIA 5062 isolated from soybean in Brazil and the type strain of *B. yuanmingense* CCBAU 10071^T, isolated from *Lespedeza cuneata* in China; however, bootstrap support was low (46 %). Subcluster I.3 included strains SEMIA 6059 from *Psophocarpus tetragonolobus* (USA), SEMIA 656 from *Neonotonia wightii* (Brazil) and SEMIA 6077 from *Stylosanthes* sp. (Brazil). Five SEMIA strains of distinct tropical legumes formed subcluster I.4: SEMIA 6028 from *Desmodium uncinatum* (Zimbabwe), three Brazilian strains (SEMIA 6146 from *Centrosema* sp., SEMIA 6148 from *N. wightii* and SEMIA 6069 from *Leucaena leucocephala*) and strain SEMIA 6093 from *Aeschynomene americana* (USA). Subcluster I.5 was formed by the Brazilian strains SEMIA 6152 from *Calopogonium* sp. and SEMIA 6160 from *Albizia lebbeck*. Seven SEMIA strains related to the type strains of *B. elkanii* and *B. pachyrhizi* (Supplementary Fig. S1) were included in subcluster I.6; the majority were symbionts of *Glycine max*, but there were also symbionts of other legumes.

Cluster II of the *nodA* tree included strain SEMIA 6319 isolated from *Arachis hypogaea* in Bolivia and *Bradyrhizobium* sp. Lcamp8 isolated from *Lupinus campestris* in Mexico (Fig. 1; Table 3). Cluster III included SEMIA 6101 from *Dalbergia nigra*, SEMIA 6440 from *Arachis pintoi* and SEMIA 6099 from *Dimorphandra exaltata*, all isolated in Brazil. Cluster IV included strain SEMIA 6374, a symbiont of *Arachis pintoi* of unknown origin, and *Bradyrhizobium* sp. CH2309, a symbiont of *Lupinus albescens* from Brazil. Cluster V included SEMIA 6156 from *Crotalaria spectabilis* (Brazil) and SEMIA 928 from *Lupinus* sp. (site of isolation unknown), in addition to *Bradyrhizobium* sp. Lpol9, Ch2355, USDA 3505, FTA7, BLUH1 and ICED, all symbionts of *Lupinus* species of distinct geographic origin (from the study of Stępkowski *et al.*, 2007). Finally, cluster VI included five SEMIA strains, all symbionts of *Acacia mearnsii*, and also *Bradyrhizobium* sp. USDA 3475 from *Acacia melanoxylon*, all isolated in Brazil. In subclusters I.4 and I.5, and in cluster III, symbionts of legumes of the subfamily Papilionoideae were interspersed with other members of the subfamilies Mimosoideae (subclusters I.5 and I.4) and Caesalpinioideae (cluster III). Cluster VI comprised predominantly symbionts of *Acacia mearnsii*,

subfamily Mimosoideae, with bootstrap support of 100 %. The other cluster and subclusters had symbionts of several species of the subfamily Papilionoideae (Fig. 1; Tables 1 and 3). Overall, the *Bradyrhizobium* partial sequences of *nodA* were genetically distant from other rhizobial *nodA* sequences (*Mesorhizobium loti* MAFF303099, *Rhizobium etli* CFN42^T, *Rhizobium* sp. NGR234, *Sinorhizobium fredii* HH103, *Sinorhizobium meliloti* 1021, *Rhizobium leguminosarum* bv. *viciae* 3841 and *Rhizobium leguminosarum* bv. *trifolii* WSM1325) included in the phylogenetic tree for comparison.

The consensus region of 402–514 bp of *nodY/K* of the same 37 strains used in the *nodA* tree was analysed and the resulting phylogenetic tree had a clear correlation with the *nodA* tree (Fig. 2; Table 3). Some type and reference strains used in the phylogenetic tree of *nodA* were not included in the *nodY/K* tree, because their sequences were not available in GenBank. Furthermore, *nodY/K* genes are found exclusively in *Bradyrhizobium*, so members of other rhizobial genera could not be included for comparison. In the *nodY/K* tree, the strains also split into six clusters, but three of them were represented by only one strain (SEMIA strains 5062, 6374 and 6319). Reference/type strains were positioned exclusively in cluster I: in subcluster I.1, seven SEMIA strains grouped with *B. japonicum* USDA 110 (bootstrap support of 83 %); and in subcluster I.5, seven other SEMIA strains grouped with *B. elkanii* USDA 76^T and USDA 94 (bootstrap support of 79 %). All strains in subcluster I.1 and five of the seven strains in subcluster I.5 were symbionts of *Glycine max*. As observed in the *nodA* tree, symbionts of legumes of the subfamily Papilionoideae were interspersed with those of the subfamilies Mimosoideae (subclusters I.4 and I.5) and Caesalpinioideae (cluster III), whereas symbionts of *Acacia mearnsii*, subfamily Mimosoideae, were all grouped in cluster VI, with bootstrap support of 99 %.

Amplification of the *nodZ* region was achieved with 35 strains using the TSnodZ primers (Table 2), and fragments of 427–433 bp were obtained and analysed. The three strains that did not amplify with the primer for the *nodD–nodB* region, SEMIA strains 6014, 6192 and 6434, amplified successfully with TSnodZ primers. In contrast, the five symbionts of *Acacia mearnsii* that clustered in the *nodA* and *nodY/K* trees did not amplify with primers for the *nodZ* region. In general, clustering in the tree built with *nodZ* was very similar to clustering in the trees built with the *nodA* and *nodY/K* genes (Fig. 3; Table 3). One exception was SEMIA 6434, which did not amplify with *nodA* and *nodY/K*, and grouped with SEMIA 6374 in cluster VI. Another exception relied on SEMIA strains 6014 and 6192, which also failed to amplify with primers for *nodA* and *nodY/K*, and occupied isolated positions in the *nodZ* tree.

Nitrogen fixation *nifH* gene

PCR amplification of the *nifH* gene resulted in fragments of 671–678 bp for all 40 SEMIA strains. As mentioned above, the PCR product of SEMIA 6434 had to be cloned before sequencing and two different fragments of similar sizes were obtained and analysed, showing similarity of 97.35 %. The phylogenetic tree built with the *nifH* gene was congruent with trees built with the *nodA* and *nodY/K* genes and strains that did not amplify for those genes occupied isolated clusters in the *nifH* tree (Fig. 4; Table 3). The five symbionts of *Acacia mearnsii* were included in cluster IX of the *nifH* tree. All *nifH* sequences of *Bradyrhizobium* SEMIA strains showed great phylogenetic distance from those of other genera, such as *Rhizobium* and *Klebsiella*, and also from the photosynthetic *Bradyrhizobium* sp. strain BTAi1.

Southern hybridization of *nodY/K*, *nodA* and *nodZ* genes

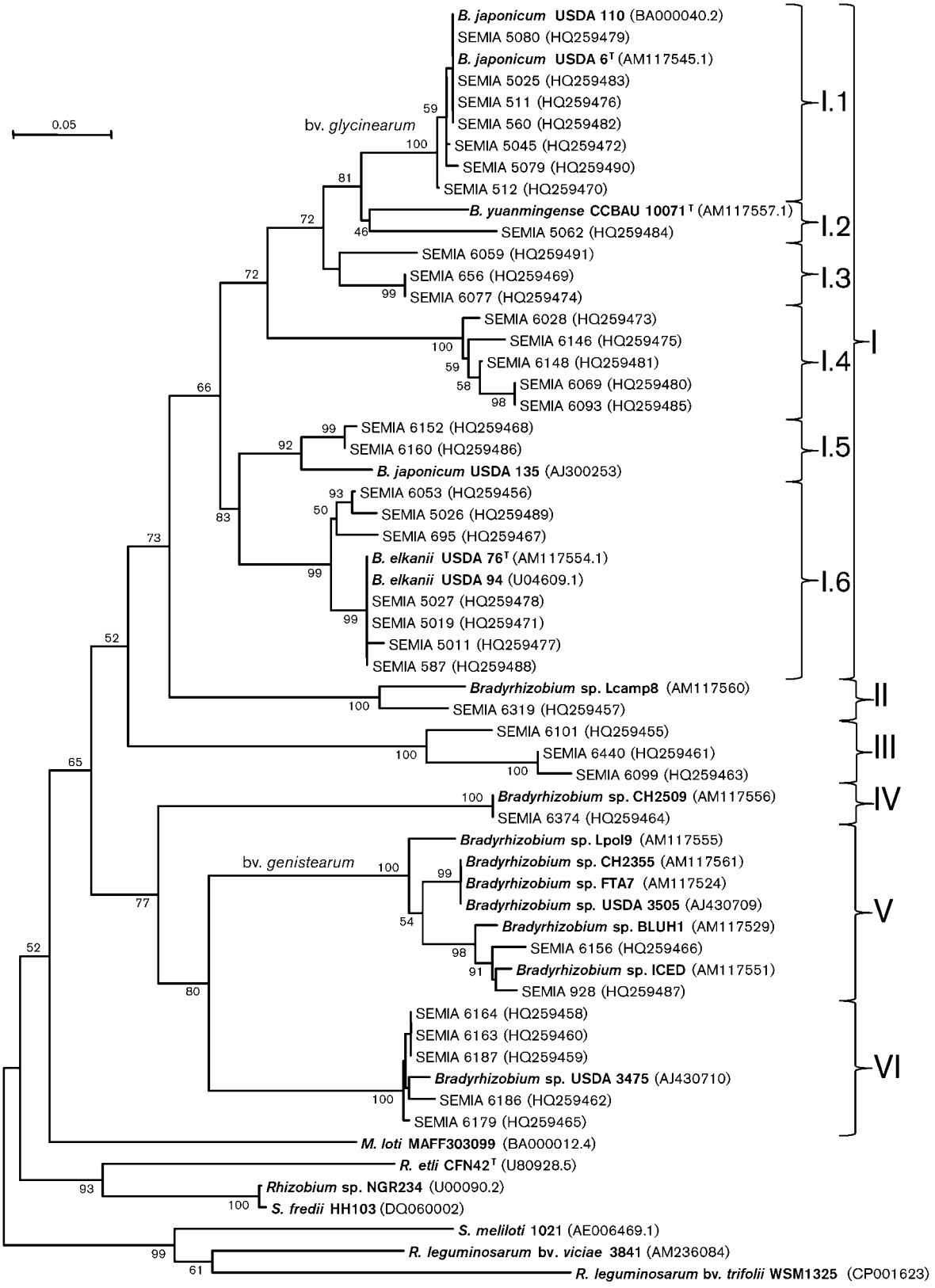
Hybridization was performed with SEMIA strains 6014, 6192 and 6434, which did not amplify with the TSnodD1-1a/TSnodB1 primers. Probes for *nodY/K* and *nodA* were prepared with *B. japonicum* USDA 6^T and *B. elkanii* USDA 76^T and no hybridization was observed for SEMIA strains. For the *nodZ* gene, probes were also prepared with strains USDA 6^T and USDA 76^T; no hybridization was observed with SEMIA strains 6163, 6164, 6179, 6186 or 6187.

Congruence of 16S rRNA and housekeeping genes with *nodA*, *nodY/K* and *nifH*

Congruence between the 16S rRNA and housekeeping genes with *nod* and *nif* trees was low (Table 3). For example, strains found in cluster I of the *nodA* tree were positioned in clusters I and II of the 16S rRNA (Table 3). The only exception observed was for strains positioned in clusters VI of *nodA* (Fig. 1), IV of *nodY/K* (Fig. 2) and IX of *nifH* (Fig. 4), that formed subclusters I.5 of the 16S rRNA and I.8 of the housekeeping trees (Supplementary Fig. S1). Interestingly, all strains belonging to these groups are symbionts of Brazilian *Acacia* species (Table 1).

DISCUSSION

Bradyrhizobium is an intriguing genus of bacteria that probably originated in tropical regions and most strains used in this study originated from legumes growing in South America, a continent that was isolated for almost 70 million years and is known to have highly diverse legume flora. Additionally, in South America, legumes showed a high level of diversity by the end of the Palaeocene period, forming one of the most species-rich groups among angiosperms in the early neotropical forest 58–60 million years ago. This might indicate that the Leguminosae have evolved in South America and/or South America was the place of their initial differentiation into three subfamilies (Sprent, 2007). Accordingly, based on ribosomal and other housekeeping genes, apparently there are also many more



varieties of *Bradyrhizobium* in tropical and subtropical regions than in temperate regions (Vinuesa *et al.*, 2008), but knowledge about nodulation and nitrogen fixation genes is still very poor. Our study, therefore, aimed to increase current knowledge by providing data for 40 bradyrhizobial strains.

Fig. 1. Phylogenetic relationships of 37 *Bradyrhizobium* SEMIA and type/reference strains based on *nodA* partial sequences. Type/reference strains are highlighted in bold. Phylogeny was inferred using the neighbour-joining method. The percentages of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. Evolutionary distances were computed using the maximum composite likelihood method and are given as the number of base substitutions per site. Positions containing gaps and missing data were eliminated from the dataset (complete deletion option). Phylogenetic analyses were conducted in MEGA4.

The hypothesis of the monophyletic origin of *nodA* in *Bradyrhizobium* has been raised from studies in which several strains fit into a well-defined phylogenetic group, in some cases showing correlation with the geographic origin of the host (Stępkowski *et al.*, 2007; Steenkamp *et al.*, 2008). Our results also support this theory, as the 37 SEMIA strains of *Bradyrhizobium* were clearly separate from members of other rhizobial genera. Additionally, studies of the evolution of the *nodA* gene in the genus *Bradyrhizobium* have indicated that the gene can be longer in this genus than in other rhizobial genera. In *Bradyrhizobium* strains isolated from various leguminous genera in several geographic regions, *nodA* coded for 209–211 amino acids, whereas in *Mesorhizobium*, *Rhizobium* and *Sinorhizobium*, the gene products were 195–198 amino acids (Moulin *et al.*, 2004; Stępkowski *et al.*, 2007). The loss of the N-terminal region of the NodA protein in other rhizobial genera suggests that the *nodA* gene may be undergoing an evolutionary progression towards a shorter sequence (Moulin *et al.*, 2004), which also gives support to the theory that nitrogen fixation in *Bradyrhizobium* could be ancestral to that in other rhizobia (Norris, 1965; Turner & Young, 2000; Lloret & Martínez-Romero, 2005). Although giving support to the monophyletic origin hypothesis, it should be noted that high diversity was detected among phylogenetic groups of *Bradyrhizobium*, therefore indicating that other strains from different geographic origins should be analysed to confirm this evolutionary hypothesis.

The previously reported correlation between *nodA* phylogenetic clusters and symbiotic host (Moulin *et al.*, 2004; Stępkowski *et al.*, 2005, 2007) was also observed in our study. In previous studies, the monophyletic *nodA* tree was split in six major branches referred to as clades I–VI. Clades I and IV comprised strains isolated from legumes native to Australia. Clade II included Genisteae and serradella isolates mainly of European origin, clade III was a large group comprising *nodA* sequences of mainly subtropical strains and clade VI included photosynthetic rhizobia. Similarly, in this study the *nodA* phylogenetic tree split the strains in distinct groups (clusters I–VI) and, in some cases, clear correlations between the clusters and the host or biogeography were observed, e.g. subcluster I.1 comprised symbionts of *Glycine max*. The presence of this specific group for *Glycine max* suggests the predominance of a specific *nodA* gene in strains related to *B. japonicum* (Supplementary Fig. S1) and the biovar *glycinearum* could be suggested for these strains. Subclusters I.2–I.6 and clusters II and III were intermixed with a variety of symbionts, but all were of pantropical origin.

Similarly to previous reports, cluster V included symbionts of *Lupinus* species (Moulin *et al.*, 2004; Stępkowski *et al.*, 2005, 2007) and for these strains the biovar *genistearum* has been suggested (Stępkowski *et al.*, 2007). However, in our study, strain SEMIA 6156, a symbiont of *Crotalaria spectabilis*, was also included in this cluster, although the ability of this strain to also nodulate *Lupinus* species has not yet been investigated.

A specific cluster was also observed for strains that were symbionts of *Acacia mearnsii* and included *Bradyrhizobium* sp. USDA 3475, a symbiont of *Acacia melanoxylon* from Brazil, positioned in cluster VI in the *nodA* tree; these strains were related to the Australian cluster (*Acacia mearnsii* is native to the temperate zone of Australia). Similar clusters were also reported by Moulin *et al.* (2004), who defined a cluster (I) that was related to the Australian group, with symbionts of *Acacia* species.

Another interesting observation is that the results from our study indicate that some strains, such as the isolates from *Arachis* species, are unique to South America, whereas others show affinity to the Australian or European clades. The peanut strains show that this promiscuous legume is nodulated by various bradyrhizobia. It is of note that none of these strains show similarity to the peanut strains from Africa described by Steenkamp *et al.* (2008). This might indicate differences between the Australian *Acacia* species, which seem to favour their 'Australian' microsymbionts, whereas peanut forms nodules with rhizobia that are available in the soil.

Clusters similar to those in the *nodA* tree were also observed in the phylogenetic tree built with *nodY/K*. Unique features within *Bradyrhizobium* are the presence of *nodY* in *B. japonicum* (420 bp) (Nieuwkoop *et al.*, 1987), *nodK* in *B. elkanii* (402 bp) (Dobert *et al.*, 1994) and *nodY/K* in *Bradyrhizobium* sp. (a symbiont of *Parasponia*, a non-legume; 411 bp) (Scott, 1986). Dobert *et al.* (1994) observed that the similarity between *nodY* of *B. japonicum* USDA 123 and *nodK* of *B. elkanii* USDA 94 was only 49 % and, in our study, the lowest similarity (36 %) was observed between SEMIA 6374 (from *Arachis pintoi*) and the other *Bradyrhizobium* strains analysed. The *nodY/K* putative gene is found downstream of *nodD1* and upstream and coregulated with the *nodABC* operon (Nieuwkoop *et al.*, 1987; Dobert *et al.*, 1994; Scott, 1986). Unfortunately,

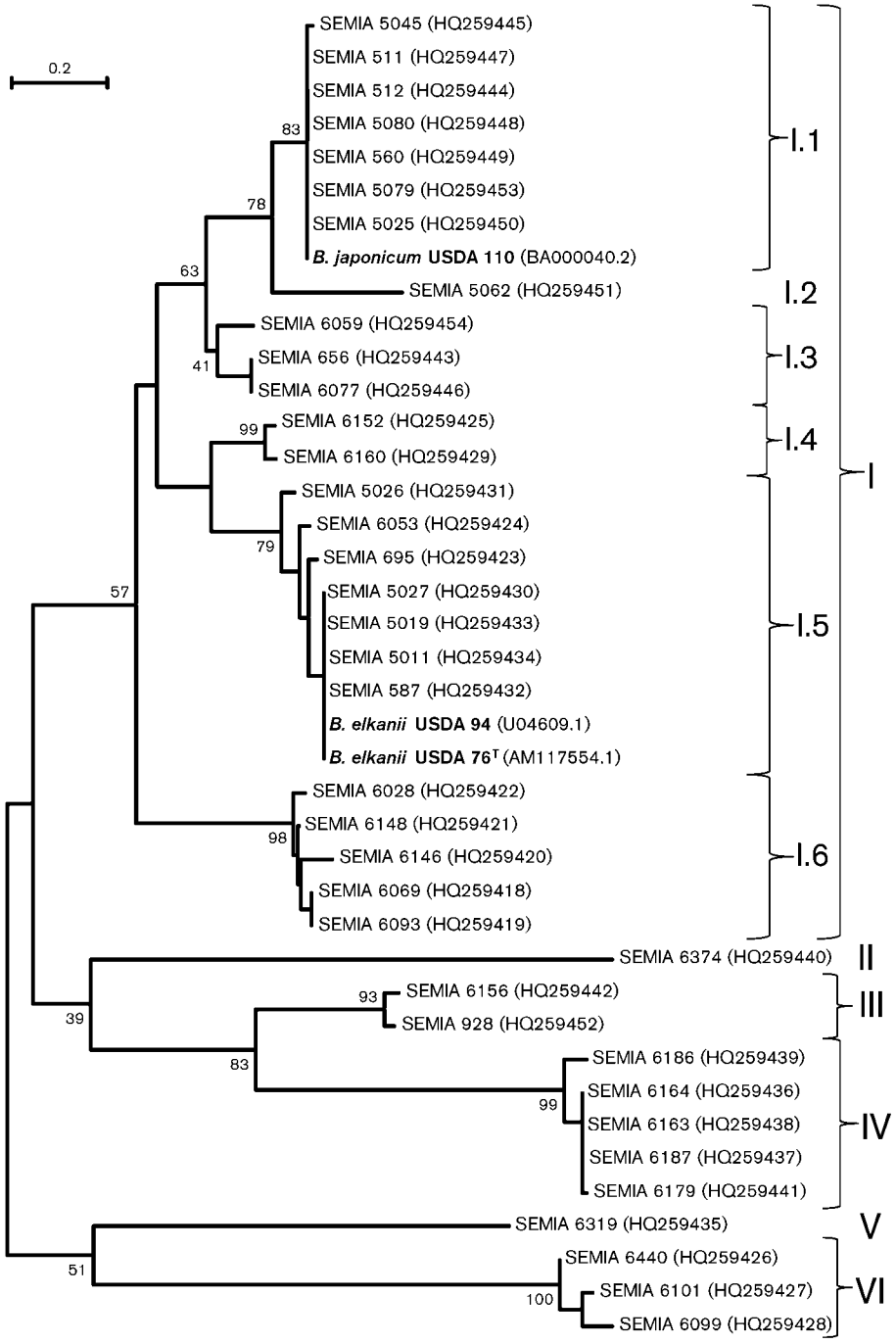


Fig. 2. Phylogenetic relationships of 37 *Bradyrhizobium* SEMIA and type/reference strains based on *nodY/K* partial sequences. See legend to Fig. 1 for further details.

despite being broadly found in a variety of *Bradyrhizobium* strains, including those reported in our study, the role of *nodY/K* is still poorly understood, with no indication that it is a protein-coding region; therefore, *nodY/K* may be regarded as a *nodD–nodA* intergenic (non-coding) region.

In a previous study, Sterner & Parker (1999) used the sequence similarity of *nodY* and *nodK* to distinguish strains of *B. japonicum* and *B. elkanii*, but a higher diversity of these genes was subsequently reported, with evidence of distinct phylogenetic groups (You *et al.*, 2002; Stępkowski *et al.*, 2003). In our study, a relationship was found between the *nodY/K* clusters and the two large 16S rRNA groups defined by Menna *et al.* (2009a), each comprising several *Bradyrhizobium* species. However, a very low

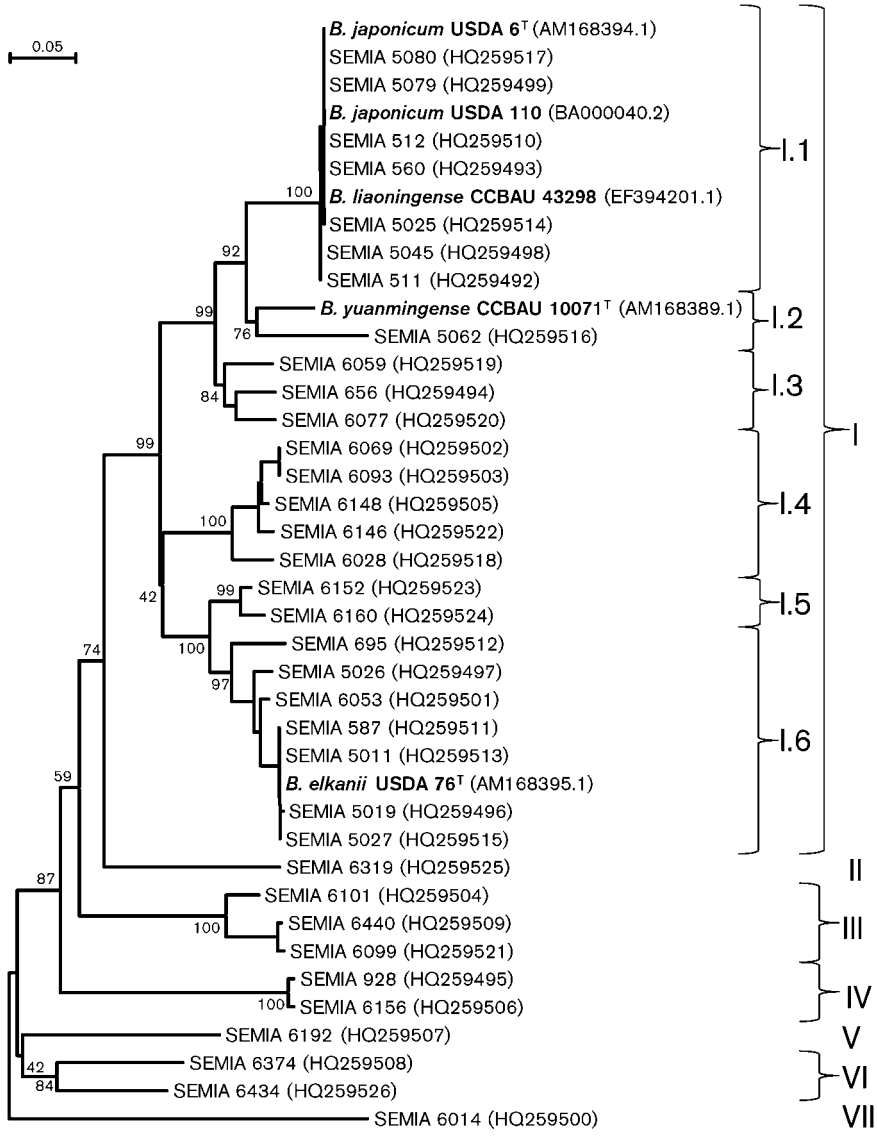


Fig. 3. Phylogenetic relationships of 35 *Bradyrhizobium* SEMIA and type/reference strains based on *nodZ* partial sequences. See legend to Fig. 1 for further details.

congruence was found between the *nodY/K* of our study and the 16S rRNA subgroups reported by Menna *et al.* (2009a). In addition, a far greater diversity was observed in the phylogeny of *nodY/K* compared to the 16S rRNA gene.

However, considering the *nodY/K* tree, several groups have shown that no clear correlation with host plant, e.g. SEMIA strains 6152 (*Calopogonium*) and 6160 (*Albizia*), symbionts from the subfamilies Papilionoideae and Mimosoideae, respectively, were grouped in subcluster I.4. However, other groups were strongly related to the symbionts, e.g. subcluster I.1 had exclusively soybean symbionts, classified as *B. japonicum* in the 16S rRNA gene tree, and cluster IV included symbionts of *Acacia mearnsii* (Mimosoideae). It is of note that a shorter fragment of *nodY/K* and *nodA*, which was about 300 bp shorter than in the other *Bradyrhizobium* strains from this study, was observed in cluster VI. The symbiosis is thought to have evolved from Caesalpinioideae to Mimosoideae and then to Papilionoideae, because nodulation frequency progresses from uncommon to very common in these subfamilies (Sprent, 2007). It is possible that the shorter *nodY/K* in *Acacia* species evolved specifically from an ancestor of the subfamily Mimosoideae and has been maintained by vertical transfer.

Neither *nodA* nor *nodY/K* was detected by amplification or hybridization in three strains: SEMIA 6014 from *Stylosanthes guianensis* (isolated in Peru), SEMIA 6192 from *Tipuana tipu* (isolated in Brazil) and SEMIA 6434 from *Inga* sp. (origin not known). Until now, the absence

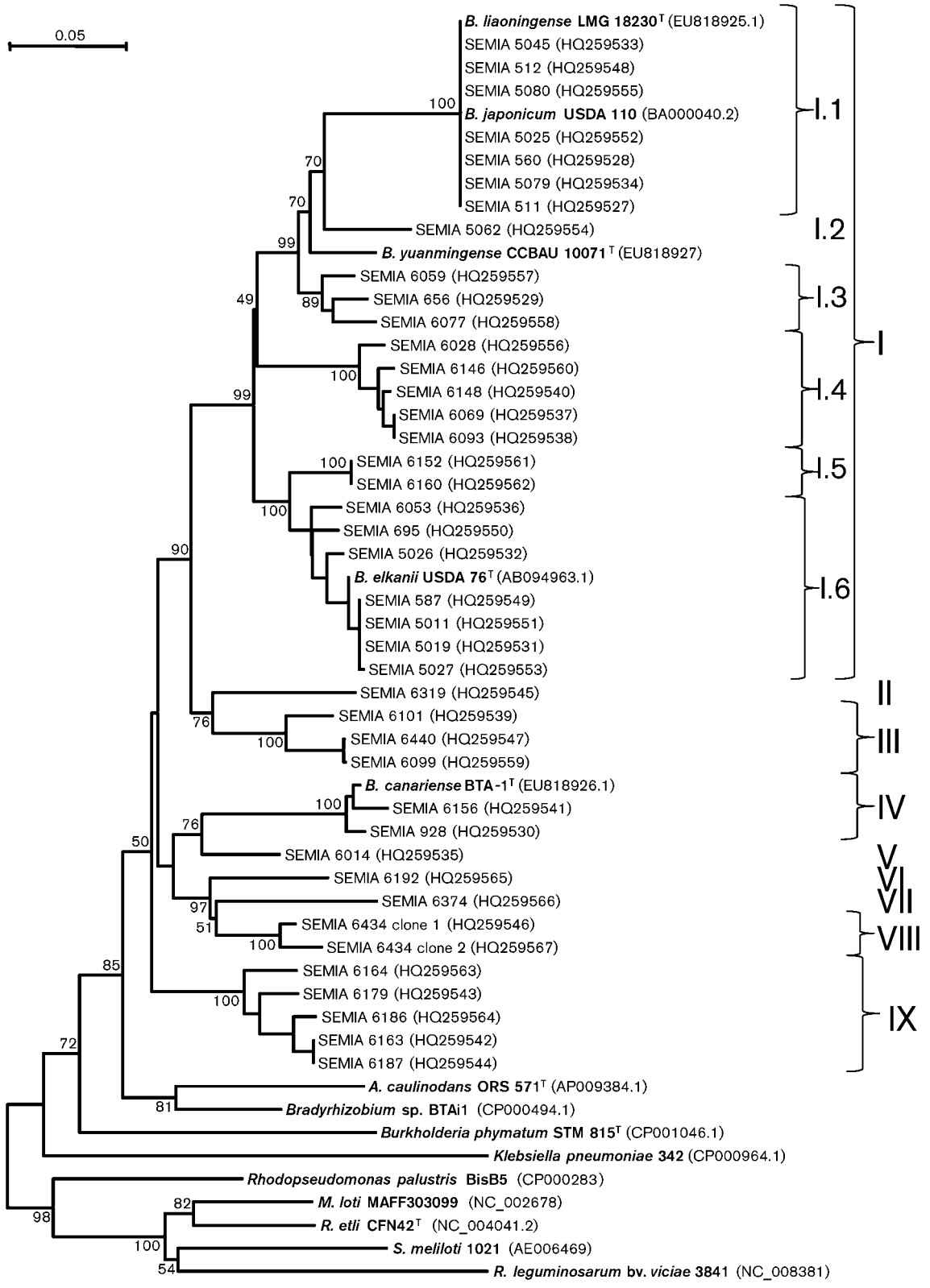


Fig. 4. Phylogenetic relationships of 40 *Bradyrhizobium* SEMIA and type/reference strains based on *nifH* partial sequences. See legend to Fig. 1 for further details.

of the common *nodABC* genes and of *nodY/K* has been reported only for photosynthetic *Bradyrhizobium* strains BTAi1 and ORS 278, symbionts from *Aeschynomene* stem nodules (Giraud *et al.*, 2007). The absence of *nodA* and *nodY/K* in these three SEMIA strains could be attributed to the specificity of the primer used in this study and the high diversity of the genes in these *Bradyrhizobium* strains. However, our results could also indicate that an alternative mechanism for initiating transcription of nodulation genes is not exclusive to the stem-nodulating group.

nodZ gene clustering has also shown high congruence with the trees built with *nodA* and *nodY/K* sequences, except that *nodZ* was present in SEMIA strains 6192, 6434 and 6014, which had no detectable *nodA* or *nodY/K*, and also that *nodZ* was absent in the group of five symbionts from *Acacia mearnsii*. *nodZ* is an unusual gene in comparison to other common nodulation genes such as *nodABC* as its expression is constitutive and independent of *nodD* (Stacey *et al.*, 1994). Chemical analysis of *nodZ* mutants of *B. japonicum* has suggested that NodZ is essential for fucosylation of the terminal reducing *N*-acetylglucosamine of the lipo-chitin oligosaccharides (LCOs) of Nod factors (Stacey *et al.*, 1994). *nodZ* was also shown to determine host specificity in a study where the transference of the *nodZ* of *B. japonicum* to *Rhizobium leguminosarum* bv. *viciae* led to the biosynthesis of LCOs fucosylated on C6 of the reducing-terminal *N*-acetylglucosamine, extending the host range to several tropical legumes, including *Macroptilium*, *Glycine*, *Vigna* and *Leucaena* (López-Lara *et al.*, 1996).

The high diversity of *nodZ* detected in our study, in addition to the variability observed in *Bradyrhizobium* from other geographic origins (Steenkamp *et al.*, 2008) gives support to the hypothesis that variability in the decorated nodulation factors might represent an important adaptation strategy, enabling nodulation of a variety of legumes. On the other hand, the absence of *nodZ* in the symbionts of *Acacia* species might indicate that a specific fucosylation of the LCOs may not be necessary for nodulation of the host plant, probably due to a close co-evolution of symbiotic partners.

In contrast to the nodulation genes, *nif* genes are found in all diazotrophic bacteria; however, it is still not clear if, from an evolutionary point of view, they are part of the symbiotic genome or the 'normal' bacterial genome (Young & Haukka, 1996). Previous studies with *Rhizobium* have shown a close phylogenetic relationship between *nifH* and 16S rRNA genes, leading to the suggestion of a common evolutionary history for both genes (Hennecke *et al.*, 1985; Young, 1992). However, this hypothesis was not supported by a further study by Eardly *et al.* (1992), suggesting that the genes located in the symbiotic plasmid may move across chromosomal backgrounds by horizontal transfer. In *B. japonicum*, the *nifH* gene and the nodulation genes are found in a symbiotic island flanked by insertion sequence elements, with a high capacity for horizontal gene transfer (Kaluza *et al.*, 1985; Kaneko *et al.*, 2002). Indeed, previous studies by our group have pointed out high rates of horizontal transfer of the symbiotic islands of *Bradyrhizobium*, even under regular field conditions (Barcellos *et al.*, 2007; Batista *et al.*, 2007). In the present study, the *nifH* tree was highly congruent with trees built with the nodulation genes *nodY/K*, *nodA* and *nodZ*, and, as in *nodA* analysis, indicates a monophyletic origin of the *nifH* gene, but which, through several episodes of vertical and horizontal gene transfer, has resulted in the high level of genetic diversity observed today.

Several studies have shown that the diversity of housekeeping genes in *Bradyrhizobium* is much higher than initially thought (e.g. Bala *et al.*, 2003; Menna *et al.*, 2009a, b), but as we pointed out before there is still little information about the diversity of nodulation and nitrogen-fixation genes within this genus. By comparing the phylogenies of nodulation and housekeeping genes of 22 *Bradyrhizobium* strains isolated from cowpea and peanut, Steenkamp *et al.* (2008) reported that overall phylogenies for the nodulation genes were incongruent with that inferred from the core genome genes, suggesting that horizontal gene transfer significantly influences the evolution of the root-nodule bacteria. Stępkowski *et al.* (2005) have also suggested that this horizontal gene transfer may be influenced by the host plant. Erratic distribution of nodulation and nitrogen-fixation genes among rhizobial species and lineages has also been highlighted (e.g. Laguerre *et al.*, 2001; Mutch *et al.*, 2003) and, altogether, the data reinforce the key role of horizontal gene transfer in genome adaptation (Ochman *et al.*, 2000; Koonin *et al.*, 2001). We believe that our study describes the highest diversity of *nod* and *nif* genes in *Bradyrhizobium* reported so far, with several well-defined groups formed with *nodY/K* and *nodA* genes and with *nifH*, in addition to a few strains occupying isolated positions. In comparison to the 16S rRNA and housekeeping gene trees, the congruence of the clusters and subclusters formed was very poor, except for groups formed for the symbionts of *Acacia mearnsii*. Three aspects of our study are particularly noteworthy: i) use of two sets of symbiotic genes, nodulation (*nodY/K*, *nodA* and *nodZ*) and nitrogen fixation (*nifH*), distantly located in the symbiosis island of *B. japonicum* (Kaneko *et al.*, 2002); ii) the congruence of all *nod* and *nif* trees found in our study; iii) the clustering of all *Bradyrhizobium* in a large group that clearly differs from other rhizobial species in the analyses of *nodA* and *nifH* (except for strain BTAi1 in the latter case). Considering these three points, our results give support to the monophyletic origin of the symbiotic island that may have spread to several *Bradyrhizobium* strains, both by vertical and horizontal gene transfer, generating a high level of diversity. From our results, we may also draw other important conclusions and hypotheses. Firstly, that the absence of *nodY/K* and *nodA* in three *Bradyrhizobium* SEMIA strains isolated from root nodules of *Stylosanthes*, *Tipuana* and *Inga* should be further investigated, as it might indicate whether the absence of the common operon *nodY/KABC* is not unique to some photosynthetic rhizobia. Another hypothesis is that the high diversity of the host-specific gene *nodZ* might indicate

Table 3. Phylogenetic grouping formed in the trees built with the following genes: 16S rRNA, housekeeping genes (Menna *et al.*, 2009a), and the genes from this study, *nodA*, *nodY/K*, *nodZ* and *nifH*

Given in the order: large 16S rRNA group, *nodA* and *nifH*. NA, Not analysed; Isol, isolated; NAmp, not amplified; ND, not defined.

Strain	16S rRNA		Housekeeping		*nodA*		*nodY/K*		*nodZ*		*nifH*	
	Cluster	Subcluster	Cluster	Subcluster	Cluster	Subcluster	Cluster	Subcluster	Cluster	Subcluster	Cluster	Subcluster
SEMIA 511	I	I.3	I	I.7	I	I.1	I	I.1	I	I.1	I	I.1
SEMIA 512	I	1.4	I	I.7	I	I.1	I	I.1	I	I.1	I	I.1
SEMIA 560	I	I.1	I	I.3	I	I.1	I	I.1	I	I.1	I	I.1
SEMIA 5025	I	I.3	I	I.1	I	I.1	I	I.1	I	I.1	I	I.1
SEMIA 5045	I	I.2	I	I.7	I	I.1	I	I.1	I	I.1	I	I.1
SEMIA 5079	I	I.4	I	I.7	I	I.1	I	I.1	I	I.1	I	I.1
SEMIA 5080	I	I.6	I	I.6	I	I.1	I	I.1	I	I.1	I	I.1
USDA 110	I	I.6	I	I.6	I	I.1	I	I.1	I	I.1	I	I.1
USDA 6^T	I	I.4	I	I.7	I	I.1	NA	NA	I	I.1	NA	NA
CCBAU 10071^T	I	I.3	I	I.2	I	I.2	NA	NA	I	I.2	Isol	Isol
SEMIA 5062	I	I.2	I	I.1	I	I.2	I	I.2	I	I.2	I	I.2
SEMIA 0656	I	I.8	I	I.4	I	I.3	I	I.3	I	I.3	I	I.3
SEMIA 6059	I	I.6	I	I.6	I	I.3	I	I.3	I	I.3	I	I.3
SEMIA 6077	I	I.8	I	I.2	I	I.3	I	I.3	I	I.3	I	I.3
SEMIA 6028	II	II.2	II	II.3	I	I.4	I	I.6	I	I.4	I	I.4
SEMIA 6069	II	II.1	II	II.3	I	I.4	I	I.6	I	I.4	I	I.4
SEMIA 6093	II	Isol	II	II.3	I	I.4	I	I.6	I	I.4	I	I.4
SEMIA 6146	II	II.2	II	II.2	I	I.4	I	I.6	I	I.4	I	I.4
SEMIA 6148	II	II.2	Isol	Isol	I	I.4	I	I.6	I	I.4	I	I.4
SEMIA 6152	II	II.1	II	II.3	I	I.5	I	I.4	I	I.5	I	I.5
SEMIA 6160	II	II.2	II	II.3	I	I.5	I	I.4	I	I.5	I	I.5
USDA 135	I	I.3	NA	NA	I	I.5	NA	NA	NA	NA	NA	NA
SEMIA 0587	II	II.2	II	II.1	I	I.6	I	I.5	I	I.6	I	I.6
SEMIA 0695	II	II.2	II	Isol	I	I.6	I	I.5	I	I.6	I	I.6
SEMIA 5011	II	II.2	II	II.1	I	I.6	I	I.5	I	I.6	I	I.6
SEMIA 5019	II	II.2	II	II.1	I	I.6	I	I.5	I	I.6	I	I.6
SEMIA 5026	II	II.2	II	II.1	I	I.6	I	I.5	I	I.6	I	I.6
SEMIA 5027	II	II.2	II	II.1	I	I.6	I	I.5	I	I.6	I	I.6
SEMIA 6053	II	II.2	II	II.3	I	I.6	I	I.5	I	I.6	I	I.6
USDA 76^T	II	II.2	II	II.1	I	I.6	I	I.5	I	I.6	I	I.6
USDA 94	NA	NA	NA	NA	I	I.6	I	I.5	NA	ND	NA	NA
SEMIA 6319	I	I.3	I	I.2	II	ND	V	ND	II	II	II	ND
NC92	NA	NA	NA	NA	II	ND	NA	NA	NA	NA	NA	NA
SEMIA 6099	II	II.2	II	II.3	III	ND	VI	ND	III	ND	III	ND
SEMIA 6101	II	II.1	II	II.2	III	ND	VI	ND	III	ND	III	ND

Table 3. cont.

Strain	16S rRNA		Housekeeping		*nodA*		*nodY/K*		*nodZ*		*nifH*	
	Cluster	Subcluster	Cluster	Subcluster	Cluster	Subcluster	Cluster	Subcluster	Cluster	Subcluster	Cluster	Subcluster
SEMIA 6440	II	II.2	II	II.3	III	ND	VI	ND	III	ND	III	ND
SEMIA 6374	I	I.8	I	I.4	IV	ND	II	ND	IV	ND	VII	ND
SEMIA 6014	I	I.7	I	I.4	NAmp	NAmp	NAmp	NAmp	VII	ND	V	ND
SEMIA 6192	I	I.8	I	I.3	NAmp	NAmp	NAmp	NAmp	V	ND	VI	ND
SEMIA 6434	I	I.8	I	Isol	NAmp	NAmp	NAmp	NAmp	IV	ND	VIII	ND
LMG 18230^T	I	I.3	I	I.1	NA	NA	NA	NA	NA	NA	I	I.1
CCBAU 43298	NA	NA	NA	NA	NA	NA	NA	ND	I	I.1	NA	ND
BTA-1^T	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	IV	ND
BC-C2^T	I	I.1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
PL7HG1^T	Isol	–	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
PAC68^T	Isol	ND	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
PAC48^T	II	II.2	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
SEMIA 0928	I	I.1	I	I.5	V	ND	III	ND	IV	ND	IV	ND
SEMIA 6156	I	I.7	I	I.3	V	ND	III	ND	IV	ND	IV	ND
USDA 3505	I	Isol	NA	NA	V	ND	NA	NA	NA	NA	NA	NA
SEMIA 6163	I	I.5	I	Isol	VI	ND	IV	ND	NAmp	NAmp	IX	I.8
SEMIA 6164	I	I.5	I	I.8	VI	ND	IV	ND	NAmp	NAmp	IX	ND
SEMIA 6179	I	I.5	I	I.8	VI	ND	IV	ND	NAmp	NAmp	IX	ND
SEMIA 6186	I	I.5	I	I.8	VI	ND	IV	ND	NAmp	NAmp	IX	ND
SEMIA 6187	I	I.5	I	I.8	VI	ND	IV	ND	NAmp	NAmp	IX	ND
USDA 3475	I	I.6	NA	NA	VI	ND	NA	NA	NA	NA	NA	NA

strategies of decoration of Nod factors to increase the host range, thus helping to explain the high diversity and poor relationship between strains assembled in cluster I (the pantropical cluster) and the host plant in the *nod* and *nifH* gene trees. Also interesting was the group of symbionts from *Acacia mearnsii*, which showed high congruence between the 16S rRNA and housekeeping genes (*dnaK*, *recA* and *glnII*) and the *nodY/K*, *nodA* and *nifH* genes, in addition to a shorter *nodY/K* and the absence of *nodZ*, strongly indicating a finely tuned co-evolution of the host plant and symbionts. The *B. japonicum* symbionts from soybean also showed high congruence of *nod* and *nif* genes with the 16S rRNA gene, albeit to a lesser extent. On the other hand, other groups deserve further study, such as the two symbionts of *Arachis* species, which occupied isolated positions in all *nod* and the *nifH* trees, with no relation with the core genes, apparently resulting from horizontal gene transfer. Therefore, despite reinforcing the theory of monophyletic origin of the symbiosis island in *Bradyrhizobium*, our study points out that events of horizontal gene transfer are common in a variety of groups, whereas, in others, vertical transfer represents the main genetic event, altogether contributing to the high level of diversity reported in this study.

ACKNOWLEDGEMENTS

The work was partially supported by CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico, Brazil), CNPq-PNPD (558455/2008-5), CNPq/MCT/MAPA (577933/2008), CPNq-Universal (470162/2009) and CNPq-Repensa (562008/2010-1). P. M. has a post-doctoral fellowship from CNPq (558455/2008-5) and M. H. is also a research fellow from CNPq. The authors thank Dr Allan R. J. Eaglesham for suggestions on the manuscript.

REFERENCES

Angus, A. A. & Hirsch, A. M. (2010). Insights into the history of the legume-betaproteobacterial symbiosis. *Mol Ecol* **19**, 28–30.

Bala, A., Murphy, P. & Giller, K. E. (2003). Distribution and diversity of rhizobia nodulating agroforestry legumes in soils from three continents in the tropics. *Mol Ecol* **12**, 917–929.

Barcellos, F. G., Menna, P., da Silva Batista, J. S. & Hungria, M. (2007). Evidence of horizontal transfer of symbiotic genes from a *Bradyrhizobium japonicum* inoculant strain to indigenous diazotrophs *Sinorhizobium* (Ensifer) *fredii* and *Bradyrhizobium elkanii* in a Brazilian Savannah soil. *Appl Environ Microbiol* **73**, 2635–2643.

Batista, J. S. S., Hungria, M., Barcellos, F. G., Ferreira, M. C. & Mendes, I. C. (2007). Variability in *Bradyrhizobium japonicum* and *B. elkanii* seven years after introduction of both the exotic micro-symbiont and the soybean host in a cerrados soil. *Microb Ecol* **53**, 270–284.

Broughton, W. J., Jabbouri, S. & Perret, X. (2000). Keys to symbiotic harmony. *J Bacteriol* **182**, 5641–5652.

Dixon, R. & Kahn, D. (2004). Genetic regulation of biological nitrogen fixation. *Nat Rev Microbiol* **2**, 621–631.

Dobert, R. C., Breil, B. T. & Triplett, E. W. (1994). DNA sequence of the common nodulation genes of *Bradyrhizobium elkanii* and their phylogenetic relationship 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 *nifH* genes. *Appl Environ Microbiol* **58**, 1809–1815.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783–791.

Freiberg, C., Fellay, R., Bairoch, A., Broughton, W. J., Rosenthal, A. & Perret, X. (1997). Molecular basis of symbiosis between *Rhizobium* and legumes. *Nature* **387**, 394–401.

Giraud, E., Moulin, L., Vallenet, D., Barbe, V., Cytryn, E., Avarre, J. C., Jaubert, M., Simon, D., Cartieaux, F. & other authors (2007). Legumes symbioses: absence of Nod genes in photosynthetic bradyrhizobia. *Science* **316**, 1307–1312.

Göttfert, M., Röthlisberger, S., Kündig, C., Beck, C., Marty, R. & Hennecke, H. (2001). Potential symbiosis-specific genes uncovered by sequencing a 410-kilobase DNA region of the *Bradyrhizobium japonicum* chromosome. *J Bacteriol* **183**, 1405–1412.

Hedges, S. B. (1992). The number of replications needed for accurate estimation of the bootstrap P value in phylogenetic studies. *Mol Biol Evol* **9**, 366–369.

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.

Kaluza, K., Hahn, M. & Hennecke, H. (1985). Repeated sequences similar to insertion elements clustered around the *nif* region of the *Rhizobium japonicum* genome. *J Bacteriol* **162**, 535–542.

Kaneko, T., Nakamura, Y., Sato, S., Minamisawa, K., Uchiumi, T., Sasamoto, S., Watanabe, A., Idesawa, K., Iriguchi, M. & other authors (2002). Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110. *DNA Res* **9**, 189–197.

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.

Koonin, E. V., Makarova, K. S. & Aravind, L. (2001). Horizontal gene transfer in prokaryotes: quantification and classification. *Annu Rev Microbiol* **55**, 709–742.

Laguerre, G., Nour, S. M., Macheret, V., Sanjuan, J., Drouin, P. & Amarger, N. (2001). Classification of rhizobia based on *nodC* and *nifH* gene analysis reveals a close phylogenetic relationship among *Phaseolus vulgaris* symbionts. *Microbiology* **147**, 981–993.

Lerouge, P., Roche, P., Faucher, C., Maillet, F., Truchet, G., Promé, J. C. & Dénarié, J. (1990). Symbiotic host-specificity of *Rhizobium meliloti* is determined by a sulphated and acylated glucosamine oligosaccharide signal. *Nature* **344**, 781–784.

Lloret, L. & Martínez-Romero, E. (2005). [Evolution and phylogeny of rhizobia]. *Rev Latinoam Microbiol* **47**, 43–60 (in Spanish).

López-Lara, I. M., Blok-Tip, L., Quinto, C., Garcia, M. L., Stacey, G., Bloemberg, G. V., Lamers, G. E., Lugtenberg, B. J., Thomas-Oates, J. E. & Spaink, H. P. (1996). NodZ of *Bradyrhizobium* extends the nodulation host range of *Rhizobium* by adding a fucosyl residue to nodulation signals. *Mol Microbiol* **21**, 397–408.

Menna, P., Hungria, M., Barcellos, F. G., Bangel, E. V., Hess, P. N. & Martínez-Romero, E. (2006). Molecular phylogeny based on the 16S rRNA gene of elite rhizobial strains used in Brazilian commercial inoculants. *Syst Appl Microbiol* **29**, 315–332.

Menna, P., Barcellos, F. G. & Hungria, M. (2009a). Phylogeny and taxonomy of a diverse collection of *Bradyrhizobium* strains based on multilocus sequence analysis of the 16S rRNA gene, ITS region and *glnII*, *recA*, *atpD* and *dnaK* genes. *Int J Syst Evol Microbiol* **59**, 2934–2950.

Menna, P., Pereira, A. A., Bangel, E. V. & Hungria, M. (2009b). rep-PCR of tropical rhizobia for strain fingerprinting, biodiversity appraisal and as a taxonomic and phylogenetic tool. *Symbiosis* **48**, 120–130.

Moulin, L., Béna, G., Boivin-Masson, C. & Stępkowski, T. (2004). Phylogenetic analyses of symbiotic nodulation genes support vertical and lateral gene co-transfer within the *Bradyrhizobium* genus. *Mol Phylogenet Evol* **30**, 720–732.

Mutch, L. A., Tamimi, S. M. & Young, J. P. W. (2003). Genotypic characterization of rhizobia nodulating *Vicia faba* from the soils of Jordan: a comparison with UK isolates. *Soil Biol Biochem* **35**, 709–714.

Nieuwkoop, A. J., Banfalvi, Z., Deshmane, N., Gerhold, D., Schell, M. G., Sirotkin, K. M. & Stacey, G. (1987). A locus encoding host range is linked to the common nodulation genes of *Bradyrhizobium japonicum*. *J Bacteriol* **169**, 2631–2638.

Norris, D. O. (1965). Acid production by *Rhizobium*: a unifying concept. *Plant Soil* **22**, 143–166.

Ochman, H., Lawrence, J. G. & Groisman, E. A. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**, 299–304.

Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**, 406–425.

Sambrook, J. & Russell, D. W. (2001). *Molecular Cloning: a Laboratory Manual*, 3rd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Scott, K. F. (1986). Conserved nodulation genes from the non-legume symbiont *Bradyrhizobium* sp. (Parasponia). *Nucleic Acids Res* **14**, 2905–2919.

Sprent, J. I. (2007). Evolving ideas of legume evolution and diversity: a taxonomic perspective on the occurrence of nodulation. *New Phytol* **174**, 11–25.

Stacey, G., Luka, S., Sanjuan, J., Banfalvi, Z., Nieuwkoop, A. J., Chun, J. Y., Forsberg, L. S. & Carlson, R. (1994). *nodZ*, a unique host-specific nodulation gene, is involved in the fucosylation of the lipooligosaccharide nodulation signal of *Bradyrhizobium japonicum*. *J Bacteriol* **176**, 620–633.

Steenkamp, E. T., Stępkowski, T., Przymusiak, A., Botha, W. J. & Law, I. J. (2008). Cowpea and peanut in southern Africa are nodulated by diverse *Bradyrhizobium* strains harboring nodulation genes that belong to the large pantropical clade common in Africa. *Mol Phylogenet Evol* **48**, 1131–1144.

Stępkowski, T., Świderska, A., Miedzinska, K., Czaplińska, M., Świderski, M., Biesiadka, J. & Legocki, A. B. (2003). Low sequence similarity and gene content of symbiotic clusters of *Bradyrhizobium* sp. WM9 (*Lupinus*) indicate early divergence of "lupin" lineage in the genus *Bradyrhizobium*. *Antonie van Leeuwenhoek* **84**, 115–124.

Stępkowski, T., Moulin, L., Krzyzańska, A., McInnes, A., Law, I. J. & Howieson, J. (2005). European origin of *Bradyrhizobium* populations infecting lupins and serradella in soils of Western Australia and South Africa. *Appl Environ Microbiol* **71**, 7041–7052.

Stępkowski, T., Hughes, C. E., Law, I. J., Markiewicz, L., Gurda, D., Chlebicka, A. & Moulin, L. (2007). Diversification of lupine *Bradyrhizobium* strains: evidence from nodulation gene trees. *Appl Environ Microbiol* **73**, 3254–3264.

Sterner, J. P. & Parker, M. A. (1999). Diversity and relationships of bradyrhizobia from *Amphicarpaea bracteata* based on partial *nod* and ribosomal sequences. *Syst Appl Microbiol* **22**, 387–392.

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

Vinuesa, P., Rojas-Jiménez, K., Contreras-Moreira, B., Mahna, S. K., Prasad, B. N., Moe, H., Selvaraju, S. B., Thierfelder, H. & Werner, D. (2008). Multilocus sequence analysis for assessment of the biogeography and evolutionary genetics of four *Bradyrhizobium* species that nodulate soybean on the Asiatic continent. *Appl Environ Microbiol* **74**, 6987–6996.

You, Z., Marutani, M. & Borthakur, D. (2002). Diversity among *Bradyrhizobium* isolates nodulating yardlong bean and sunnhemp in Guam. *J Appl Microbiol* **93**, 577–584.

Young, J. P. W. (1992). Phylogenetic classification of nitrogen-fixing organisms. In *Biological Nitrogen Fixation*, pp. 43–79. Edited by G. Stacey, H. R. Burris & H. J. Evans. New York: Chapman and Hall.

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

Zehr, J. P., Wyman, M., Miller, V., Duguay, L. & Capone, D. G. (1993). Modification of the Fe protein of nitrogenase in natural populations of *Trichodesmium thiebautii*. *Appl Environ Microbiol* **59**, 669–676.

Zhao, C. T., Wang, E. T., Chen, W. F. & Chen, W. X. (2008). Diverse genomic species and evidences of symbiotic gene lateral transfer detected among the rhizobia associated with *Astragalus* species grown in the temperate regions of China. *FEMS Microbiol Lett* **286**, 263–273.