Available online at www.sciencedirect.com
SCIENCE DIRECT®
Systematic and Applied Microbiology 28 (2005) 702–716
SYSTEMATIC AND APPLIED MICROBIOLOGY
www.elsevier.de/syapm

Molecular systematics of rhizobia based on maximum likelihood and Bayesian phylogenies inferred from *rrs*, *atpD*, *recA* and *nifH* sequences, and their use in the classification of *Sesbania* microsymbionts from Venezuelan wetlands

Pablo Vinuesa[a,*], Claudia Silva[a], María José Lorite[b], María Luisa Izaguirre-Mayoral[c], Eulogio J. Bedmar[b], Esperanza Martínez-Romero[a]

[a]*Centro de Ciencias Genómicas, UNAM Apdo 565A, Cuernavaca, Morelos, México*
[b]*Departamento de Microbiología del Suelo y Sistemas Simbióticos. Estación Experimental del Zaidín, CSIC, Granada, Spain*
[c]*Instituto Venezolano de Investigaciones Científicas, Centro de Microbiología y Biología Celular, Laboratorio de Biotecnología y Virología Vegetal, Caracas, Venezuela*

Received 10 April 2005

Abstract

A well-resolved rhizobial species phylogeny with 51 haplotypes was inferred from a combined *atpD*+*recA* data set using Bayesian inference with best-fit, gene-specific substitution models. Relatively dense taxon sampling for the genera *Rhizobium* and *Mesorhizobium* was achieved by generating *atpD* and *recA* sequences for six type and 24 reference strains not previously available in GenBank. This phylogeny was used to classify nine nodule isolates from *Sesbania exasperata*, *S. punicea* and *S. sericea* plants native to seasonally flooded areas of Venezuela, and compared with a PCR-RFLP analysis of *rrs* plus *rrl* genes and large maximum likelihood *rrs* and *nifH* phylogenies. We show that *rrs* phylogenies are particularly sensitive to strain choice due to the high levels of sequence mosaicism found at this locus. All analyses consistently identified the *Sesbania* isolates as *Mesorhizobium plurifarium* or *Rhizobium huautlense*. Host range experiments on ten legume species coupled with plasmid profiling uncovered potential novel biovarieties of both species. This study demonstrates the wide geographic and environmental distribution of *M. plurifarium*, that *R. galegae* and *R. huautlense* are sister lineages, and the synonymy of *R. gallicum*, *R. mongolense* and *R. yanglingense*. Complex and diverse phylogeographic, inheritance and host-association patterns were found for the symbiotic *nifH* locus. The results and the analytical approaches used herein are discussed in the context of rhizobial taxonomy and molecular systematics.

Keywords: Bayesian phylogenetics; Maximum likelihood; *Mesorhizobium plurifarium*; *Rhizobium huautlense*; *Sesbania*; Nitrogen-fixation; Plasmids; Molecular systematics; Recombination

*Corresponding author. Tel.: +52 777 3131697; fax: +52 777 3175581.
E-mail address: vinuesa@ccg.unam.mx (P. Vinuesa).

0723-2020/$ - see front matter © 2005 Elsevier GmbH. All rights reserved.
doi:10.1016/j.syapm.2005.05.007

Introduction

The molecular systematics of rhizobia is largely based and revised in the light of *rrs* phylogenies [33,55,56]. They are equated with species phylogenies and play therefore a key role in species identification and taxonomy. However, there are several caveats with *rrs*-based taxonomy, particularly when dealing with closely related species and genera. First, it is well established that a single gene phylogeny should not be confounded with a species phylogeny, as each gene tree is the random outcome of a unique genealogical process [10,23,45,46]. Second, ribosomal operons are highly prone to homologous recombination, often resulting in a mosaic structure that causes problems for phylogeny estimation using traditional tree reconstruction methods [42,50], as illustrated herein. Finally, in several genera the level of *rrs* sequence polymorphism is too low to permit the resolution of statistically significant sequence (species) splits [17,43].

When recombination and lateral transfer events are shaping gene genealogies, several unlinked loci and individuals per species should be analyzed to approximate species phylogenies [8,23,43,45,46]. This is particularly so when only short stretches of individual loci are sequenced. We present a well-resolved rhizobial species phylogeny based on combined *atpD*+*recA* sequences and show its utility as a molecular systematic framework in which novel sequences can be inserted for evolutionary, taxonomic and ecological investigations, as compared with a *rrs*-based phylogeny. We took advantage of the existing sequence databases, which are extended herein with 9 *Sesbania* nodule isolates, six type and 15 reference strains in the genera *Rhizobium* and *Mesorhizobium*. Therefore a reasonably dense taxon sampling was obtained for these genera. Maximum likelihood and Bayesian statistical frameworks were used for model selection, phylogeny estimation and hypothesis testing. The merits of the phylogenetic frameworks provided by the protein-coding loci and *rrs* sequences were evaluated in relation to their relative topological stability and resolving power. They were further used to classify nine nodule isolates from different *Sesbania* species native to Venezuelan wetlands.

Sesbania punicea (Cav) Benth, *S. sericea* (Willd.) Link, and *S. exasperata* (Kunth) are flood-tolerant shrubs that exhibit rapid growth during the rainy season. *S. sericea* is widely distributed in Venezuela, whereas the other two species are confined to sites in the northeast of the country. Recent studies have shown that tropical wetlands inherently have a low N-content due to the strongly leached soils brought about by seasonal flooding, along with the low N mineralization and high denitrification rates prevalent in such environments [1]. Other reports have demonstrated the almost ubiquitous N_2-fixation capacity of diverse herbaceous and tree legumes native to different tropical wetlands, and showed that this process contributes significantly to the N input in such ecosystems [16,18]. *Sesbania* species are widespread and commonly found in tropical wetlands of both hemispheres, either as native or cultivated plants. Reports from tropical Africa, Asia, Puerto Rico, Central and South America indicate that *Sesbania* species can be effectively nodulated by *Bradyrhizobium* sp. [5], *Rhizobium huautlense* [49], *R. gallicum* and *R. tropici* [57], *Sinorhizobium saheli* and *S. terangae* [20], *Azorhizobium caulinodans* [6], *Mesorhizobium* sp. and *M. plurifarium* [47,48].

Presently, no information is available on the rhizobial strains harboured by native *Sesbania* species of Venezuelan wetlands, or on their symbiotic effectiveness. Therefore, additional objectives of this study were (i) to gain basic knowledge on the diversity and taxonomic identity of rhizobia associated with the native *Sesbania* species mentioned above, (ii) to determine key symbiotic properties of these bacteria such as host range and their effect on total N content in shoots, (iii) to evaluate the suitability of combined *rrs*+*rrl* PCR-RFLPs for rapid and accurate strain identification at the species level.

Finally, inferences about the phylogeography and inheritance patterns exhibited by the symbiotic *nifH* locus of these bacteria were made from a maximum likelihood phylogeny of forty eight *nifH* haplotypes representative of the currently reported sequence diversity of this locus for the genera *Rhizobium*, *Sinorhizobium* and *Mesorhizobium*.

Materials and methods

Bacterial isolation and basic microbiological characterization

Field-grown *S. punicea* (Cav.) Benth, *S. exasperata* (Kunth) and *S. sericea* (Willd.) Link plants were collected at the localities of Rio Negro, San Casimiro and Charallave in the Aragua State of Venenzuela, respectively. All plants were at the vegetative stage of growth and were harvested during the rainy season. They were classified by botanists from the Botanical Garden of the Central University of Caracas. Nodules of these species were exclusively found on the underwater roots and were embedded within a thick white aerenchyma (Izaguirre-Mayoral, unpublished observations). The root nodule endosymbionts were isolated and purified on solid (1.5% agar) yeast extract-mannitol mediumYEM [37]. Doubling times were calculated from the exponential growth phase of cultures grown in YEM broth. The native isolates from *S. punicea*, *S. exasperata*

and *S. sericea* were denoted as Sp, Se and Ss, respectively.

Plant material and growth conditions

Seeds from field-grown and mature *S. punicea*, *S. exasperata*, and *S. sericea* were surface-sterilized in a 3% sodium hypochlorite solution for 10 min, and sown in Leonard jars [15]. Uninoculated plants served as controls. Plants were grown in controlled-environment chambers for 25 days at 12/12 h light/dark cycle, 27/24 °C, 60% relative humidity and a photosynthetic photon flux density (400–700 nm) of 620 µmol photon $m^{-2} s^{-1}$, emitted by fluorescent lamps at the CHBC-Caracas facilities. To study the capacity of native isolates to form stem nodules, pre-germinated seeds of *S. rostrata*, *S. punicea*, *S. exasperata*, and *S. sericea* were placed in 20 × 3 cm sterilized test tubes containing 5 cm of 1.5% agar–N-free Fahraeus nutrient solution. Nutrient solution was added up to the base of the stem. For stem inoculation, the liquid inoculum (2×10^9 cells ml^{-1}) was rubbed onto the non-flooded area of the stem with a sterilized cotton tip. Total N content determination and statistical analysis of the growth parameters were performed using one-way analysis of variance, as described previously [15].

Isolation of genomic DNA and plasmid profiling

Genomic DNA of the native isolates was prepared with a CTAB-based protocol as described elsewhere [44]. The plasmid content of the isolates was analyzed in 0.7% agarose gels by a modified Eckhardt procedure [7].

PCR-RFLP analysis of the *rrn* operon

The SSU rRNA gene region (*rrs*) was amplified using primers fD1and rD1 [52]. A 2.2 kb fragment of the *rrl* (23S rDNA) gene was amplified with primers P3 and P4 [41], using the same cycling parameters and reaction mixture as for the *rrs* amplification reported previously [44]. The *rrs* amplification products were digested with *Cfo*I, *Dde*I, and *Msp*I, and the *rrl* amplicons with *Cfo*I, *Hae*III and *Hin*fI (Amersham and New England Biolabs, UK). The restriction patterns were resolved electrophoretically on 2% (w/v) MetaPhor agarose gels (Biozym, Hess. Oldendorf, Germany), as described [44].

Computer-assisted analysis of electrophoretic patterns

Comparative analysis of PCR-RFLPs was performed with GelCompar*II* V.2 (Applied Maths BVBA, Belgium) as described [44], based on binary band-matching tables from which pair-wise similarity matrices were calculated using Dice's similarity coefficient [4], and used for cluster analysis with the neighbor-joining method [31]. Nodal support was inferred by non-parametric bootstrap analysis with 1000 pseudoreplicates [9]. This was performed in GelCompar by randomly resampling characters (presence–absence of bands) with replacement from the binary band-matching table and calculation of the corresponding Dice coefficients.

Amplification and sequencing of *rrs*, *atpD*, *recA* and *nifH* loci

Nearly full length SSU rRNA sequences were obtained as described [43]. Partial *atpD*, *recA* and *nifH* gene fragments were amplified with primer pairs (*atpD*255F/*atpD*754R, *recA*41F/*recA*640R and *nifH*40F/*nifH*817R; the coordinates correspond to *B. japonicum* USDA110 gene positions) and amplification protocols reported elsewhere [46]. Amplification products were purified and subjected to cycle sequencing as described previously [46].

Evolutionary analyses of nucleotide sequences

Multiple nucleotide sequence alignments were generated by ClustalW, as implemented in DAMBE [53], guiding the alignment of coding sequences by the corresponding protein alignments. The *rrs* sequences were aligned using the Ribosomal Database Project-II [21]. Multiple alignments were scanned for intragenic recombination using Bootscanning [32], Chimaera [27] and GENECONV1.81 [34], as implemented in RDP2 [22] and explained in more detail previously [46]. Phylogenetic trees were inferred under the maximum likelihood optimality criterion using PhyML2.4.3 [12] and the nucleotide substitution model selected by the Akaike information criterion (AIC), as implemented in MODELTEST3.06 [26] and described in detail elsewhere [25,46]. Among-site rate variation was approximated by a discrete gamma distribution with eight rate categories [54], each category being represented by its mean. The robustness of ML topologies was inferred by non-parametric bootstrap tests [9] using PhyML and 100 data pseudoreplicates. The Shimodaira–Hasegawa test [36] was used for hypothesis testing in a ML framework as implemented in PAUP*4b10 [39] under the RELL model with 10 000 replicates.

Bayesian phylogenetic analyses [14] of concatenated sequence alignments using mixed models were performed with MrBayes3b4 [30]. Characters within combined sequence sets were partitioned by gene, as described previously [46], using the substitution model selected by the AIC. The different data partitions were allowed to evolve at different rates, but branch lengths

were assumed to be proportional across partitions using a rate multiplier [30]. All substitution model parameters were also allowed to be independent across partitions. Default MrBayes priors on model parameters were used, except for Ratepr = variable (for multiple partition analyses). Metropolis-coupled Markov chain Monte Carlo (MCMCMC) was used to estimate the posterior probability distribution using three incrementally heated chains with the temperature parameter set to 0.15, all four chains starting from different random trees. Each analysis was replicated twice for 2×10^6 generations, sampling the posterior distribution every 200th generation. Evidence for convergence of the different Markov chains was obtained by examining the correlation between the posterior probabilities of individual clades in pairwise comparisons among runs, and by comparing their generation plots for overall model likelihood (marginal $-\ln L$ against generations) and all model parameters. We checked that all runs had similar mean and variance of model likelihood and parameter values after discarding the burn-in trees [14], and pooled the samples from the stationary phases of the two independent runs (with burnin set to 50% of sampled trees) to obtain a 50% majority rule consensus tree using MrBayes, as explained previously [46].

Accession numbers for nucleotide sequences

The 71 nucleotide sequences generated in this study (4 *rrs*, 30 *atpD*, 30 *recA* and 7 *nifH*) are deposited in GenBank under the acc. nos. AY688589–AY688623, AY907351-AY907378, AY907427, AY907434, AY907436, AY907441, AY907449, AY907452, AY907456 and AY907460. They are individually specified on the phylograms shown in Figs. 3, 4 and 6.

Results

Symbiotic efficiency and host range of Venezuelan *Sesbania* spp. isolates

Three isolates were obtained from the nodules of native *S. exasperata* (Se), *S. punicea* (Sp) and *S. sericea* (Ss) plants respectively (Table 1). They were fast-growers, with a mean generation time of 2–4 h.

All isolates formed effective and determinate (lacking a persistent apical meristem) root nodules when inoculated on their original hosts. The only significantly superior combination in terms of shoot mass production was the symbiosis between Ss140 and *S. sericea* (Table 1), whereas the worst combination was that of Sp45 with *S. punicea*; the other combinations appeared to be equally efficient. None of the isolates formed nodules on *Glycine max*, *Stylosanthes guianensis*, *Medicago sativa, Clitoria alternata* or *Aeschynomene* spp. Nodulation results for the other hosts tested are shown in Table 1. Notably, the Se and Ss isolates displayed the broadest range of efficient symbiotic interactions, whereas the Sp isolates nodulated efficiently only their original hosts and *Vigna unguiculata*. However, the Se isolates did not nodulate the latter host or *Leucaena leucocephala* plants. Non-inoculated control plants did not survive the seedling stage, dying after the abscission of cotyledons.

All Venezuelan nodule isolates, as well as *A. caulinodans* strain ORS571^T [6], formed effective root nodules when inoculated on *S. rostrata* (Table 2).

Table 1. Classification, plasmid profiles, symbiotic efficiency and host range of the Venezuelan *Sesbania* spp. nodule isolates

Sesbania species[a]	Isolate	Classification[b]	Plasmid profile[c]	Shoot weight[d] (mg dry wt pl^{-1})	Host range[e]				
					Vu	*Pv*	*Ll*	*Ma*	*Cp*
S. punicea	Sp42	*M. plurifarium*	None	29.0±1.7 b	E	NE	NE	NE	NE
	Sp45	*M. plurifarium*	P2/500	26.2±1.7 c	E	NE	NE	NE	NE
	Sp53	*M. plurifarium*	P4/450	28.9±1.8 b	E	NE	NE	NE	NE
S. sericea	Ss121	*R. huautlense*	P3/900, 300, 25	28.3±1.7 b	E	E	E	E	Nd
	Ss140	*M. plurifarium*	P2/500	32.0±2.1 a	E	E	E	E	Nd
	Ss145	*R. huautlense*	P1/900, 400, 50	28.4±2.0 b	E	E	E	E	Nd
S. exasperata	Se127	*R. huautlense*	P1/900, 400, 50	29.5±2.0 ab	—	E	—	E	Nd
	Se128	*R. huautlense*	P1/900, 400, 50	28.2±2.1 b	—	E	—	E	Nd
	Se130	*R. huautlense*	P1/900, 400, 50	29.6±2.0 ab	—	E	—	E	Nd

[a]Original hosts of the nodule isolates.
[b]Based on the combined results of *rrs*+*rrl* PCR-RFLPs, *rrs* and *atpD*+*recA* phylogenies.
[c]Plasmid profile type (P1–P4) and estimated sizes of the plasmid bands in kilobases.
[d]Results are the mean of four replicates using the original hosts of the isolates. Numbers in column followed by the same letter are not statistically different at $p<0.05$. Plants were grown for 25 days.
[e]Hosts are *Vu* = *Vigna unguiculata*; *Pv* = *Phaseolus vulgaris*; *Ll* = *Leucaena leucocephala*, *Ma* = *Macroptilium atropurpureum*; *Cp* = *Centrosema pubescens*. E = effective; NE = non-effective; — = not nodulated; Nd = not determined.

Table 2. Root and stem nodulation, shoot and nodule mass, and total nitrogen content in the shoots of *S. rostrata* plants inoculated with Venezuelan rhizobial isolates and *A. caulinodans* ORS571^T

Isolate[a]	Root nodules	Stem nodules	Shoot weight[b] (mg dry wt pl^{-1})	Root nodule mass[b] (mg dry wt pl^{-1})	Total nitrogen[b] (mg N g^{-1})
Mp Sp42	+	−	187bc	29c	14c
Mp Sp45	+	−	129e	36b	16b
Mp Sp53	+	−	162d	35b	16b
Rh Ss121	+	+	190b	24d	16b
Mp Ss140	+	−	173c	17e	14c
Rh Se127	+	+	203b	25d	17b
Rh Se128	+	+	190b	16e	15c
Rh Se130	+	+	132e	5f	13c
ORS571^T	+	+	629a	68a	43a

+ = presence of nodules, − = absence of nodules.

[a]The species assignations of the isolates are indicated as *Mp* for *M. plurifarium* and *Rh* for *R. huautlense*. The host species of original isolation are denoted by Se = *Sesbania exasperata*; Sp = *S. punicea*; Ss = *S. sericea*, respectively.

[b]Results are the mean of four replicates. Numbers in columns followed by the same letter are not statistically different at $p \leqslant 0.05$. Plants were grown for 25 days in 20 × 3 cm tubes containing medium-logged agar as the rooting substrate.

Stem nodules (actually on lateral root primordia) were detected on *S. rostrata* plants inoculated with strain ORS571^T and with the *R. huautlense* isolates Ss121, Se127, Se128 and Se130. None of the *M. plurifarium* isolates formed nodules on the stems of *S. rostrata* (Table 2). Plants inoculated with the Se, Ss and Sp isolates displayed significantly smaller aerial and root nodule mass as well as lower shoot nitrogen content, than the plants inoculated with *A. caulinodans* ORS571^T (Table 2). The low shoot masses and total N levels recorded are due to the constrained growth of the plants in the test tubes used for these assays. None of the Se, Sp and Ss isolates, nor *A. caulinodans* ORS571^T, formed stem nodules on the three native Venezuelan *Sesbania* species.

Analysis of combined 16S + 23S rDNA restriction patterns

The *rrs* and *rrl* PCR-amplifications yielded single bands of approximately 1.5 and 2.2 kb respectively. A Dice/NJ analysis of the combined *rrs* + *rrl* PCR-RFLPs is shown in Fig. 1. All three Se strains and isolates Ss121 and Ss145 were grouped with high (83%) bootstrap support (BS) in a cluster that contained the *R. huautlense* type strain. All Sp isolates and Ss140 were grouped in a weakly supported cluster (62% BS) that included the type strains of *M. amorphae* and *M. plurifarium*. Bootstrap analysis resolved very well the *Mesorhizobium* spp. strain cluster (99% BS) but not the *Rhizobium* spp. cluster (Fig. 1).

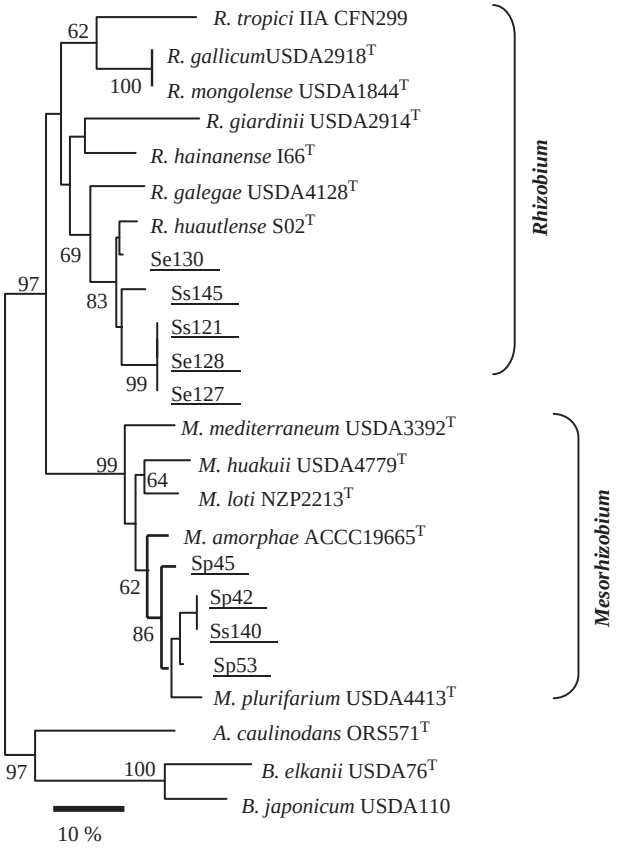

Fig. 1. Dice (S_D) neighbor-joining analysis of combined *rrs* + *rrl* PCR-RFLPs using three restriction enzymes to digest each amplification product, showing the genetic affinities of the nine Venezuelan *Sesbania* isolates (underlined) with respect to selected reference strains. Those highlighted with a superscript T are type strains. Bootstrap support values out of 1000 pseudoreplicates of the data set are provided as percentages at the corresponding nodes when > 50%. The bar indicates 10% distance, calculated as 1—S_D.

Plasmid profiling

Four plasmid profiles (P) could be discerned among the nine Venezuelan *Sesbania* spp. isolates (Table 1 and Fig. 2). The most abundant profile (P1) was shared by

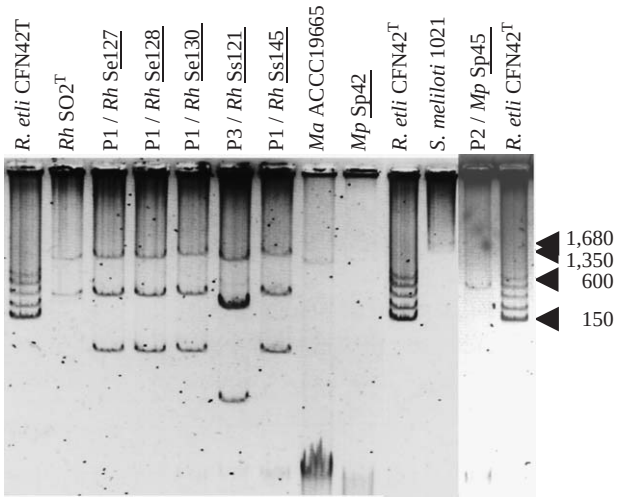

Fig. 2. Plasmid profiles of the nine Venezuelan *Sesbania* isolates and reference strains, as resolved by Eckhardt gel electrophoresis. Size estimates for each plasmid for every profile (P1–P4) are provided in Table 1. *Rh* = *R. huautlense*, *Ma* = *M. amorphae*.

the three Se isolates and strain Ss145, all classified as *R. huautlense*. Profile P1 was very similar to that of *R. huautlense* S02^T, except for the presence of an additional small plasmid (Fig. 2). The *R. huautlense* strain Ss121 displayed a distinct and unique profile (P3) with three plasmid bands, the central one possibly representing a doublet. Two profiles (P2 and P4) were found among the *M. plurifarium* isolates, both consisting of a single plasmid, whereas no plasmids could be observed in strain Sp42 (Table 1 and Fig. 2). In summary, the Venezuelan *R. huautlense* isolates harboured three (or more) plasmids, whereas in *M. plurifarium* strains a single or no plasmid was observed.

Maximum likelihood *rrs* gene phylogeny

Fig. 3 shows the phylogenetic placement of four Venezuelan *Sesbania* isolates (Se128, Se130, Sp42 and Sp45) in the context of a maximum likelihood (ML) topology including forty nine *rrs* haplotypes from the type strains of most of the species currently described in the genera *Rhizobium*, *Mesorhizobium* and *Sinorhizobium*. Some additional sequences closely related to those from *M. plurifarium* and *R. huautlense* were included to get a better delineation of these particular clades. Basic alignment characteristics are listed in Table 3.

The sequences from the *Sesbania exasperata* isolates Se128 and Se130 were recovered in a highly supported (100% BS) clade along with those from the *R. huautlense* and *R. galegae* type strains and *Rhizobium* sp. OK-55, which is congruent with their grouping in the the *rrs*+*rrl* PCR-RFLP analysis (compare with Fig. 1). The Sp42 and Sp45 sequences were recovered in a weakly supported (70% BS) clade along with those from three *M. plurifarium* strains isolated in Mexico or Africa, including that of LMG11892^T, the type strain. This result is also consistent with the PCR-RFLP analysis (Fig. 1). However, the sequences of strains Ls38 and Sn2, also classified as *M. plurifarium* [47], were not recovered in the previous clade.

At a higher taxonomic level, the generic clades grouping *Mesorhizobium* and *Sinorhizobium rrs* sequences are well (94% BS) to moderately supported (80% BS), respectively, both having a similar level of nucleotide diversity and no clear internal subdivisions. The *Rhizobium-Agrobacterium* clade is also well supported (90% BS), but has a significantly higher nucleotide diversity (about 2.5 times higher) than that found in the *Sinorhizobium* or *Mesorhizobium* clades. Some internal structure is apparent in the *Rhizobium* clade, the sequences related to those of *R. gallicum*, *R. leguminosarum* and *R. galegae* forming a moderately supported cluster (87% BS), those from *R. giardinii*, *R. daejeonense*, *R. rubi*, *R. radiobacter*, *R. vitis* and *R. undicola* representing independent lineages with uncertain relationships among them. These sequences were found to contain significant internal mosaicism as illustrated in Figs. 4D and E, which contributes to the poor phylogenetic resolution found among most bipartitions in that group and other clades. In particular, the sequences from *S. morelense* LMG20571 (AJ42077) and *Ensifer adhaerens* LMG20582 (AJ420775) were excluded from the analyses because their presence strongly distorted the topology of the *Rhizobium* clade, making it paraphyletic (data not shown). This is most likely due to the mosaic structure of their *rrs* sequences, which contain segments similar to *Rhizobium* while others are highly similar to *Mesorhizobium* sequences, as shown in Figs. 4A–C.

Maximum likelihood and Bayesian species phylogenies of the *Rhizobiaceae* and *Phyllobacteriaceae* inferred from combined *atpD*+*recA* sequences

The Bayesian phylogeny shown in Fig. 5A adds 30 new *atpD* and *recA* sequences to those reported in previous papers [11,46,51]. Among them are those corresponding to the 9 Venezuelan *Sesbania* isolates (underlined), those for the type strains of *R. huautlense* S02^T, *M. amorphae* ACCC19665^T and *M. plurifarium* USDA4413^T, *R. gallicum* R602^T, *Rhizobium mongolense* USDA1844^T, and *R. yanglingense* SH22623^T, and 15 reference strains. Our *atpD* and *recA* sequences for the type strain of *M. amorphae* ACCC19665^T and *M. plurifarium* USDA4413^T are different from those reported by Weir and colleagues for *M. amorphae* ICMP15022^T and for the *recA* sequence of *M. plurifarium* ICMP13640^T [51].

The *atpD* and *recA* partitions were combined since the Bayesian majority-rule consesus trees inferred for each of them resulted in non-significant Shimodaira–Hasegawa test values when compared with the topology inferred from the combined partitions (Table 4). When directly compared to each other, the *atpD* and *recA*

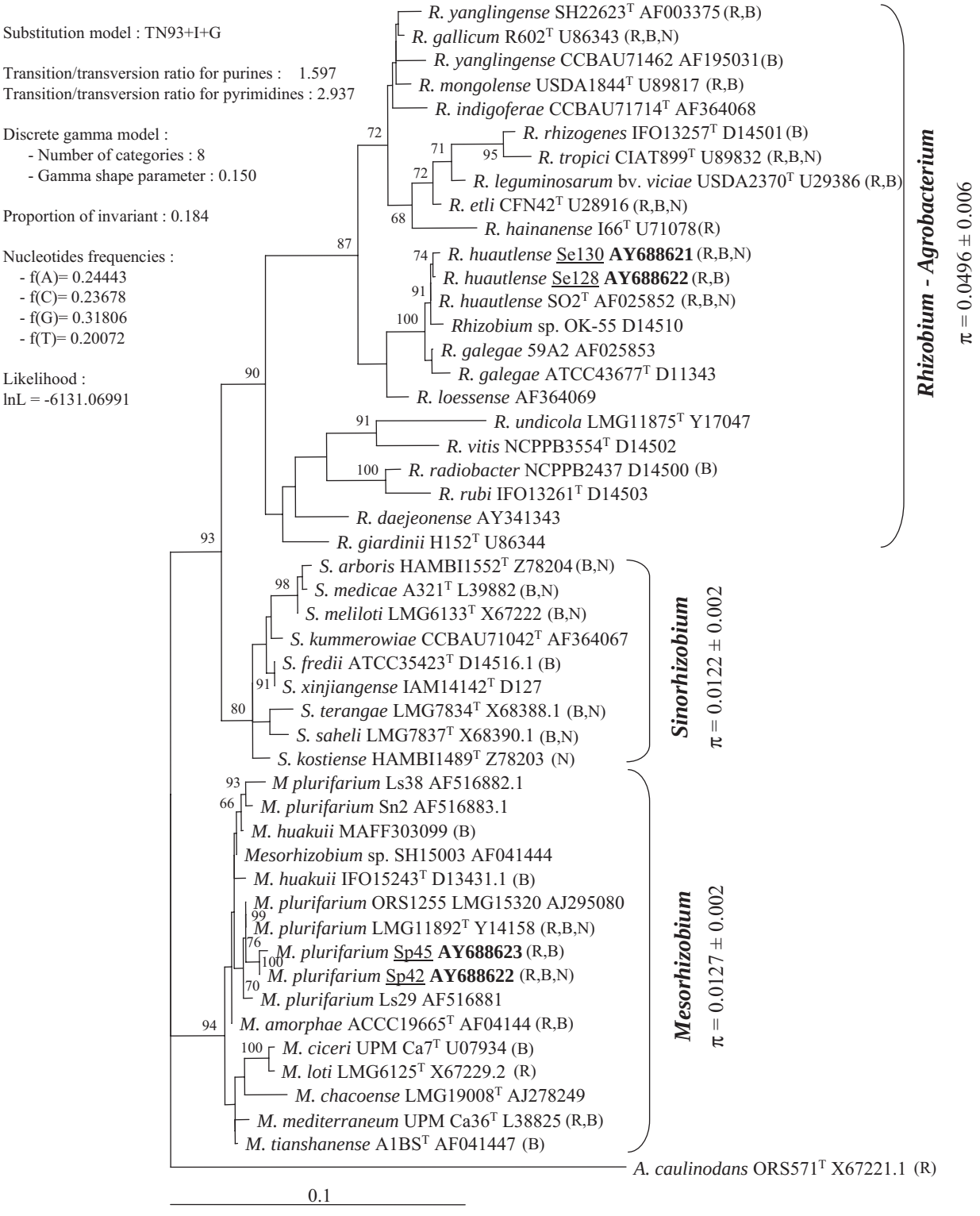

Fig. 3. Maximum likelihood phylogram based on nearly full-length *rrs* (16S rDNA) sequences of 49 rhizobial strains, including four Venezuelan *Sesbania* isolates (underlined) recovered under the best-fit model shown. Maximum likelihood bootstrap support values for 100 pseudoreplicates of the data set are provided at the corresponding nodes. Accession numbers for the sequences generated in this work are shown in bold face. The codes after the acc. nos. indicate strains used in other analyses (R = RFLP, B = Bayesian phylogeny of *atpD+recA* sequences, N = ML *nifH* phylogeny). π refers to nucleotide diversity estimates using Tamura-Nei distances. Basic alignment characteristics are listed in Table 3.

partitions yielded significant values (Table 4), indicating the presence of conflicting signals. However, the 95% credible set of trees for the combined data contained about 5.9–6.1 times fewer topologies than those obtained for the individual *atpD* and *recA* partitions (Fig. 5F), demonstrating in rigorous statistical terms the adequacy of combining both data sets.

The model selected by the AIC for the separate and combined *atpD* and *recA* partitions was GTR + I + G in all cases. Basic alignment characteristics for each gene are listed in Table 3.

Since the focus of this study is on the *Mesorhizobium* and *Rhizobium* clades, the phylogeny was rooted with three *Bradyrhizobium* sequences. Two major perfectly monophyletic ingroup clades are recognized, the *Mesorhizobium* and the *Rhizobium-Sinorhizobium* lineages.

Several species relationships within the *Mesorhizobium* clade were not resolved in the ML analysis. However, that corresponding to *M. plurifarium*, which contains four Venezuelan isolates, the type strain USDA4413^T (isolated from *Acacia senegal* in Senegal), and two additional reference strains is strongly supported (PP = 1; 89% BS).

The *Sinorhizobium* clade is strongly supported and splits off at a basal position with respect to the *Rhizobium* clade, both genera clearly sharing a recent common evolutionary ancestor.

Four highly supported species lineages (I–IV in Fig. 5A) were resolved within the *Rhizobium* clade. However, the relationships among them are not resolved. The mean pairwise between-lineage genetic distances for lineages I–IV and the *Sinorhizobium* clade (lineage V) are shown in Fig. 5D. Notice that the mean between-lineage genetic distances for species grouped in the genus *Rhizobium* are smaller than that between the *Rhizobium* and *Sinorhizobium* clades (Fig. 5E), indicating that a small but significant genetic differentiation is found between both genera with the sequences used. However, a more exhaustive locus and taxon sampling, particularly of sinorhizobia, would be required to make a more robust estimate of the genetic differentiation between these genera.

Table 3. Basic sequence alignment characteristics

Loci	No. seq[a]	No. aln sites[b]	No. gaps	No. C sites[c]	No. V sites[d]	No./% Pi sites[e]
rrs	49	1422	64	1087	323	258/18
atpD	32	453	12	265	188	167/37
recA	32	480	0	271	209	197/41
nifH	48	506	3	252	252	209/41

[a]Number haplotypes (unique sequences).
[b]Number of aligned sites.
[c]Number of constant sites.
[d]Number of variable sites.
[e]Number and percentage of parsimony informative sites.

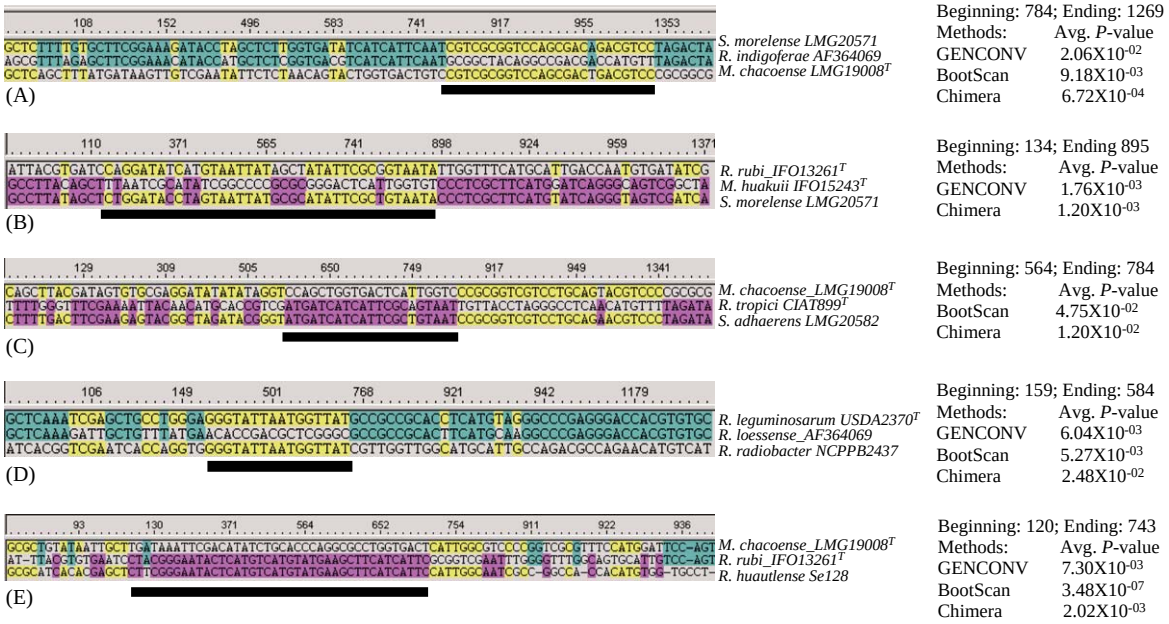

Fig. 4. Evidence for intragenic mosaicism in some of the *rrs* sequences used to reconstruct the phylogeny shown in Fig. 3, as identified by GENCONV, Chimera and BootScan. Panels A, B and C show sequences containing probable recombinant sequences derived from different genera. Panels D and E show examples of gene mosaicism resulting from recombination between congeneric species. Only significant ($P < 0.05$) predictions made by at least two independent programs (using the Bonferroni correction for multiple comparisons) are presented. The predicted recombinant segments are indicated by a solid bar. Their positions (coordinates) in the corresponding alignemts are provided. Acc. nos. for the *S. morelense* and *S. adhaerens* strains are AJ420776 and AJ420775, respectively. For the other strains the acc. nos. are those given in Fig. 3.

Lineage I groups *R. huautlense* (containing five of the Venezuelan isolates) and *R. galegae* as sister species, lineages, which is also consistent with the result from PCR-RFLP (Fig. 1) and *rrs* sequence analyses (Fig. 3). Lineage II tightly groups *R. gallicum*, *R. mongolense* and *R. yanglingense* strains from China, France and Mexico as a monophyletic lineage with very low genetic diversity. Lineage III groups *R. tropici* with *bona fide* agrobacteria such as *Agrobacterium tumefaciens* and *A. rhizogenes*, which is clearly in conflict with their grouping in the *rrs* phylogeny (Fig. 3). Lineage IV groups *R. etli* and *R. leguminosarum* strains. The type strain of *R. etli* (CFN42^T) is placed as an outlayer of this lineage (also on the trees inferred from single partitions), whereas the other well characterized *R. etli* strains from different geographic regions form a well supported sister group to *R. leguminosarum* strains. It is worth noting that the *R. etli*-*R. leguminosarum* clade has no sister relationship to the *R. gallicum*-*R. mongolense* clade, as suggested by our *rrs*-based phylogeny (Fig. 3) and other published *rrs* trees [33].

Compelling evidence for the convergence and adequate mixing of the Markov chains [14] is shown by the overlapping overall model likelihood (marginal –ln *L*) generation plots for two independent runs started from random topologies (Figs. 5B), and the high correlation found between the posterior clade probabilities in pairwise comparisons among runs (Fig. 5C).

Evolutionary inferences from a maximum likelihood *nifH* gene phylogeny

Fig. 6 shows a ML phylogram for 48 *nifH* haplotypes, 7 of which are reported in this study. Basic alignment characteristics are shown in Table 3. Nine major sequence clusters of *Mesorhizobium*, *Rhizobium* and *Sinorhizobium* are resolved with ⩾97% BS (labeled I–IX) when the phylogeny is rooted with outgroup sequences from *Rhodobacter capsulatus*, *Azotobacter chroococcum* and *Klebsiella pneumoniae*. The 45 ingroup sequences form a strongly supported clade that reveals their common (monophyletic) evolutionary origin. The relationships among the nine ingroup clades are not resolved. All, but clade I, group only congeneric sequences, although many of them group different species, highlighting frequent lateral transfer of the locus within genera. The boxed area of clade I (subclade I-D) groups *Rhizobium huautlense* and *Sinorhizobium* spp. *nifH* sequences from isolates recovered from different *Sesbania* species growing in different continents, highlighting a possible case of transfer of the locus across genera. However, not all *Sesbania* isolates contain *nifH* sequences grouped in the I-D clade. The S42, S53 and S140 sequences were recovered in clade VII, which contains exclusively *M. plurifarium* isolates, regardless of the host or geographic origin. Therefore, some sequence clades (i.e. I-A, I-B, I-C) seem to correlate with broad geographic subdivisions as previously suggested by Haukka et al. [13], but this does not hold for most of them (i.e. I-D, IV, V, VI, VII). Similarly, some (i.e. I-D, III, V, VIII, IX), but not all clades correlate well with particular host genera or tribes. Altogether, the *nifH* gene phylogeny reveals a complex mode of inheritance, with vertical and horizontal components, as well as contrasting phylogeographic patterns and diverse degrees of host coevolution, as found in other studies [13,19,43,46].

Discussion

Our typing and phylogenetic analyses strongly support the classification of the nine Venezuelan *Sesbania* spp. isolates analyzed herein as *M. plurifarium* [2] or

Fig. 5. Bayesian phylogeny inferred from 51 combined and congruent *atpD*+*recA* sequences (panel A), including the nine Venezuelan *Sesbania* isolates (underlined) and key reference strains from the families *Rhizobiaceae* and *Phyllobacteriaceae*. The topology represents the 50% majority rule consensus tree derived from 9998 (4999×2) trees resulting from the convergent and pooled post burn-in samples (burnin = 5001 = 50%) obtained from 2 independent MCMCMC samplers, run for 2×10^6 generations and sampled every 200th, using gene-specific substitution model (GTR + I + G) parameters. Clades resolved with significant posterior probability (⩾0.95) are indicated. Significant ML-bootstrap support values (obtained under the GTR + I + G model) are also provided when ⩾70%. Clades significantly resolved only in the Bayesian analysis are marked with an asterisk. Nodes significantly resolved in both the Bayesian and ML analyses (⩾0.95/70%, respectively) of the individual *atpD* and *recA* partitions are indicated by up- and downward pointing arrow heads, respectively. The acc. nos. for the sequences generated in this study are highlighted in bold face. Those from *R. leguminosarum* 3841, *B. japonicum* USDA110 and *M. huakuii* MAFF303099 were retrieved from the corresponding genome sequence databases. The remaining ones were generated by Gaunt et al. [11] and Vinuesa et al. [46]. The codes after the acc. nos. indicate strains used in other analyses (R = RFLP, r = ML *rrs* phylogeny, N = ML *nifH* phylogeny). Panel B shows the superimposed marginal –lnL generation plots for the two MCMCMC runs. Panel C shows the correlation plot for clade posterior probabilities of trees randomly sampled from runs 1 and 2 and their correlation value. Panel D shows the pairwaise between-group average distances for lineages I–V, while panel E shows the mean between-group genetic distances for *Rhizobium*/*Agrobacterium lineages* (I–IV) and that for the pooled *Rhizobium*/*Agrobacterium* lineages vs. the *Sinorhizobium* (V) strains. Panel F presents a comparision of the phylogenetic uncertainty present in the different data sets as measured by the number of topologies present in the corresponding Bayesian 95% credible set of trees (pooled samples from two MCMCMC samplers as described above).

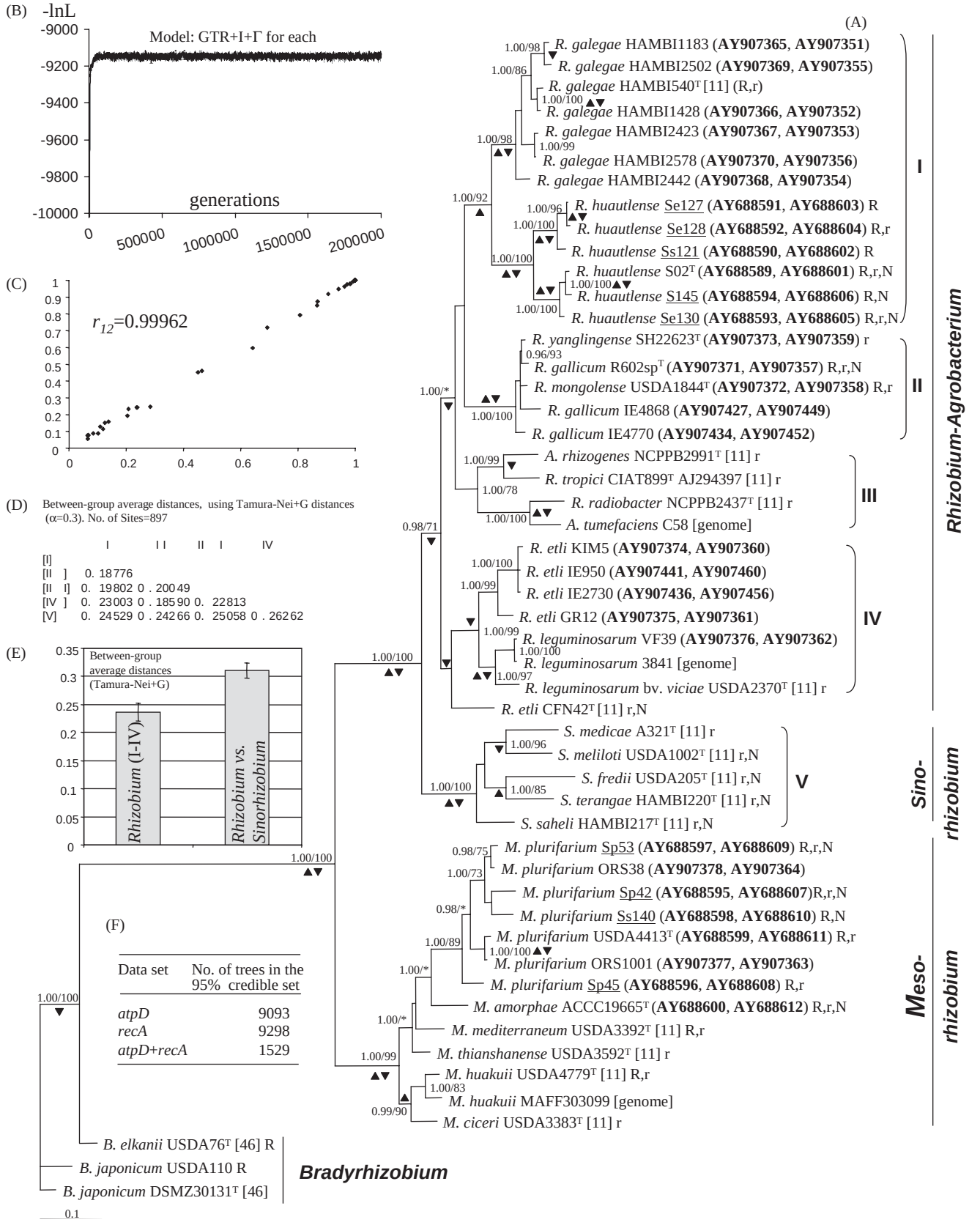
(B) -lnL
Model: GTR+I+Γ for each
generations
(C)
r_{12}=0.99962
(D) Between-group average distances, using Tamura-Nei+G distances (α=0.3). No. of Sites=897
I II II I IV
[I]
[II] 0. 18776
[II I] 0. 19802 0 . 20049
[IV] 0. 23003 0 . 18590 0. 22813
[V] 0. 24529 0 . 24266 0. 25058 0 . 26262
(E)
Between-group average distances (Tamura-Nei+G)
Rhizobium (I-IV)
Rhizobium vs. Sinorhizobium
(F)
Data set | No. of trees in the 95% credible set
atpD 9093
recA 9298
atpD+recA 1529
(A)
R. galegae HAMBI1183 (AY907365, AY907351)
R. galegae HAMBI2502 (AY907369, AY907355)
R. galegae HAMBI540T [11] (R,r)
R. galegae HAMBI1428 (AY907366, AY907352)
R. galegae HAMBI2423 (AY907367, AY907353)
R. galegae HAMBI2578 (AY907370, AY907356)
R. galegae HAMBI2442 (AY907368, AY907354)
R. huautlense Se127 (AY688591, AY688603) R
R. huautlense Se128 (AY688592, AY688604) R,r
R. huautlense Ss121 (AY688590, AY688602) R
R. huautlense S02T (AY688589, AY688601) R,r,N
R. huautlense S145 (AY688594, AY688606) R,N
R. huautlense Se130 (AY688593, AY688605) R,r,N
R. yanglingense SH22623T (AY907373, AY907359) r
R. gallicum R602spT (AY907371, AY907357) R,r,N
R. mongolense USDA1844T (AY907372, AY907358) R,r
R. gallicum IE4868 (AY907427, AY907449)
R. gallicum IE4770 (AY907434, AY907452)
A. rhizogenes NCPPB2991T [11] r
R. tropici CIAT899T AJ294397 [11] r
R. radiobacter NCPPB2437T [11] r
A. tumefaciens C58 [genome]
R. etli KIM5 (AY907374, AY907360)
R. etli IE950 (AY907441, AY907460)
R. etli IE2730 (AY907436, AY907456)
R. etli GR12 (AY907375, AY907361)
R. leguminosarum VF39 (AY907376, AY907362)
R. leguminosarum 3841 [genome]
R. leguminosarum bv. viciae USDA2370T [11] r
R. etli CFN42T [11] r,N
S. medicae A321T [11] r
S. meliloti USDA1002T [11] r,N
S. fredii USDA205T [11] r,N
S. terangae HAMBI220T [11] r,N
S. saheli HAMBI217T [11] r,N
M. plurifarium Sp53 (AY688597, AY688609) R,r,N
M. plurifarium ORS38 (AY907378, AY907364)
M. plurifarium Sp42 (AY688595, AY688607)R,r,N
M. plurifarium Ss140 (AY688598, AY688610) R,N
M. plurifarium USDA4413T (AY688599, AY688611) R,r
M. plurifarium ORS1001 (AY907377, AY907363)
M. plurifarium Sp45 (AY688596, AY688608) R,r
M. amorphae ACCC19665T (AY688600, AY688612) R,r,N
M. mediterraneum USDA3392T [11] R,r
M. thianshanense USDA3592T [11] r
M. huakuii USDA4779T [11] R,r
M. huakuii MAFF303099 [genome]
M. ciceri USDA3383T [11] r
B. elkanii USDA76T [46] R
B. japonicum USDA110 R
B. japonicum DSMZ30131T [46]
Bradyrhizobium
I
II
III
IV
V
Rhizobium-Agrobacterium
Sino-rhizobium
Meso-rhizobium

Table 4. Pairwise Shimodaira–Hasegawa tests

Tree[a]	*atpD*[b] $\Delta \ln L/P$	*recA*[b] $\Delta \ln L/P$
AtpD+recA	4.473/0.58	21.614/0.21
atpD	—	163.696/0.0001
recA	160.627/0.0001	—

[a]Bayesian 50% majority rule consensus topologies inferred under the best-fit model selected by the Akaike Information Criterion (GRT + I + G in all cases), after pooling the convergent samples of two independent runs and discarding the first 50% of trees sampled from each Markov chain as burn-in trees.

[b]Refers to the *atpD* or *recA* alignments and GTR + I + G parameters optimized under the maximum-likelihood criterion in PAUP*, given the corresponding Bayesian 50% majority rule consensus topologies. The first digit in each column corresponds to the difference in log likelihood between topologies and the second one is the *P* value of the SH test (significant if <0.05).

R. huautlense [49]. We showed that the analysis of combined *rrs*+*rrl* PCR-RFLPs [29,41] is an efficient tool for the rapid and accurate identification of *Sesbania* isolates at the species level, as they were recovered in well supported clusters that were consistent with the corresponding sequence-based phylogenetic clades.

Interestingly, although *Mesorhizobium* sp. strains were reported to be outcompeted by *R. huautlense* under flooding conditions in Mexican soils [48], this appears not to be the case in the naturally flooded areas of Venezuela, as both species were recovered from *S. sericea*, indicating that not only *R. huautlense*, but also *M. plurifarium*, has ecotypes adapted to aquatic environments. The host range of the Venezuelan *M. plurifarium* isolates contrasts markedly with that reported by de Lajudie et al. [3] for Brazilian, Senegalese and Sudanese isolates, none of which nodulated *S. rostrata*, *S. grandiflora*, *S. pubescens* nor *Phaseolus vulgaris*. Therefore, our data indicate that different biovars or ecotypes of *M. plurifarium* are found in different ecosystems, and corroborate that this species has a broad host range and a wide geographic and environmental distribution in the tropics and subtropics [3,13,28]. It is worth noting that strains Sp45, Sp53 and Ss140 each had a single plasmid, although the species has previously been described as lacking plasmids [47].

The *nifH* sequences of the Venezuelan *M. plurifarium* isolates group in a well resolved clade along with other *Mesorhizobium* strains isolated from divergent hosts in the subfamilies Caesalpinioideae (*Chamaechrista*) and Mimosoideae (*Acacia* and *Leucaena*). The reference strains in this clade have contrasting geographic origins (Kenya, Sudan, Mexico and Brazil). They were previously studied by Haukka et al. [13] and classified as *M. plurifarium* by de Lajudie et al. [3]. This finding indicates that *M. plurifarium* strains form a well delineated lineage of widely distributed broad host range strains that share monophyletic *atpD*, *recA* and *nifH* alleles, allowing robust species identification using a combination of these molecular markers.

In contrast, *R. huautlense* strains seem to have a narrower ecological range and distribution. These were previously known to occur only in wetlands of central Mexico [48,49]. In this work we show that *R. huautlense* is also native to seasonally flooded areas of Venezuela, hence lending support to the notion that this species is well adapted to water-logged soils [48]. Although the three *S. exasperata* isolates had a plasmid pattern similar to that of the *R. huautlense* type strain [49], the host range of the Venezuelan isolates differed markedly from the Mexican strains, particularly in that the former were not able to nodulate *Leucaena leucocephala* plants, but nodulated *P. vulgaris* effectively. The *R. huautlense* isolates from *S. sericea* showed a broader symbiotic effectiveness, both in terms of host range and N_2-fixation capacity. This observations again suggest the possible existence of several biovarieties in this species. Sequencing of nodulation genes in the future may provide phylogenetic evidence for the existence of such biovarieties as reported recently for *Bradyrhizobium* strains [43,46]. A noteworthy feature of the *nifH* clades I-D and VII is that they group haplotypes with a nearly global distribution in diverse tropical and subtropical ecosystems, corresponding to nodule endosymbionts from the genera *Rhizobium*, *Sinorhizobium* and *Mesorhizobium* nodulating native legumes. This distribution pattern most likely reflects a cosmopolitan distribution of these particular *nifH* alleles.

The discussion above on the ecology phylogeography and symbiotic features of the novel *M. plurifarium* and *R. huautlense* strains rests upon a correct species assignation of these isolates. One convenient approach to infer species phylogenies is to combine sequences of several unlinked and neutrally evolving loci which are shown a priori to yield congruent topologies by appropriate statistical tests [8,11,46]. Recombination events within and across rhizobial species are invariably found when large sequence sets are analyzed [42,46]. These represent a primary source for significant topological incongruence among different partitions, or among different regions of a particular locus. Analyses to detect recombinant sequences (intra- or intergenic mosaics) should be routinely performed when phylogenies are to be estimated using standard tree-reconstruction methods, as they all assume that a single evolutionary path generated the observed data. Failure to identify xenologous sequences can severely distort phylogenies, leading researchers to reach wrong or poor conclusions [35,42]. This was particularly evident with the *rrs* data set, where a significant number of sequences seemed to be mosaics resulting from interspecies recombination events. Consequently, rhizobial phylogenies inferred from ribosomal sequences are notoriously sensitive to strain choice. This is the likely explanation for

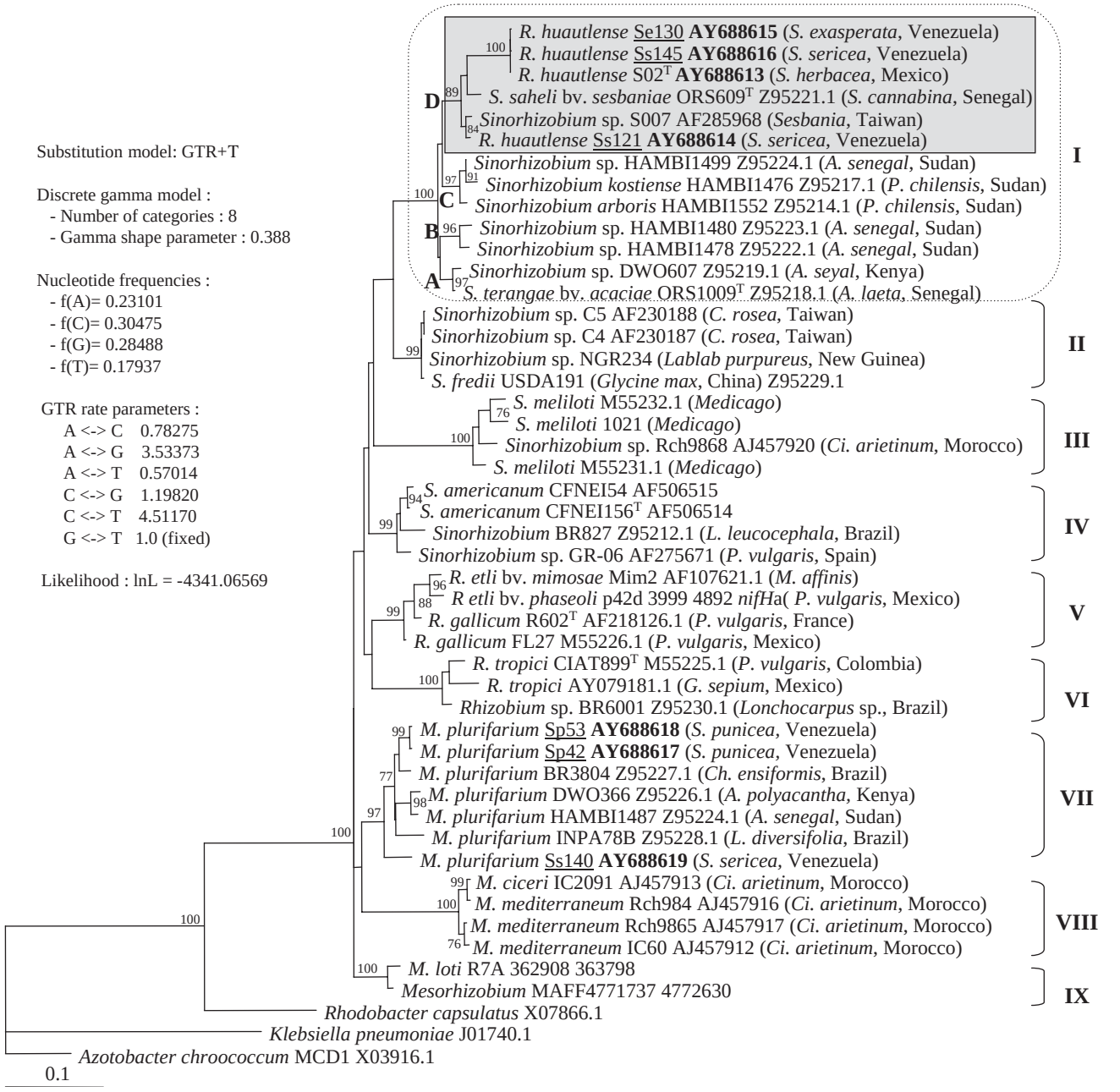

Fig. 6. Maximum likelihood phylogram depicting the evolutionary relationships among 48 *nifH* haplotypes, including six of the Venezuelan *Sesbania* isolates (underlined). The GTR + G model and indicated parameter values were used to infer the phylogeny. Bootstrap support values for 100 pseudoreplicates of the data set are provided at the corresponding nodes. Accession numbers for each sequence are provided, those generated in this study being highlighted in bold face. The host and geographic origin for each isolate are provided when known. Brackets to the right highlight nine well-resolved subclades.

some of the contradictory topologies published throughout the taxonomic literature of rhizobia, although different alignments, inadequate model choice and poor reconstruction methods also contribute to this problem. In multilocus comparisons, the identification and removal of sequences subjected to lateral transfer within and across species is also critical to infer well-resolved species phylogenies [8,46]. However, these can be easily recognized when multiple strains per taxon are analyzed. The relatively short sequences generated for the *atpD* and *recA* loci limited the power of detection methods to identify intragenic sequence mosaicism [27]. However, the *atpD* + *recA* topology was much less sensitive to taxon choice than the *rrs* tree, which is consistent with lower levels of gene mosaicism in the former loci.

In conclusion, we showed the power of multilocus sequence analyses in rhizobial molecular systematics when they are combined with dense taxon sampling and

advanced phylogenetic inference methods. The Bayesian approach was particularly powerful to estimate a well-resolved phylogenetic hypothesis for species identification. Current software implementations [30] are well suited to deal with complex inference problems such as those involving different sequence partitions [24,46]. This, along with the inherent property of Bayesian MCMC-based samplers to accommodate phylogenetic uncertainty [14], allowed the estimation of a topology that reconciled reasonably well the conflicting phylogenetic signals present in the individual *atpD* and *recA* partitions, reinforcing the congruent ones. Consequently, the tree inferred from the combined data set was better resolved and had about six times lower uncertainty than the ones estimated for each partition separately. However, Bayesian methods are known to have a tendency to produce clade overcredibility, especially when oversimplified models are used [38]. To minimize this potential pitfall, we used the complex models selected by the AIC [25] for all inferences, and evaluated topological support also by the much more conservative non-parametric bootstrap approach in a ML setting. Interesting molecular systematic aspects were highlighted by the resulting *atpD*+*recA* phylogeny: (1) the monophyly and low genetic diversity found for the *R. gallicum*-*R. mongolense*-*R. yanglingense* strains suggests that the latter two species are synonyms of *R. gallicum*, as found previously by *rrs* [56] and ITS [40] sequence analyses; (2) it confirms the apparent sister relationships between the *R. galegae*-*R. huautlense* [49] and the *R. etli*-*R. leguminosarum* lineages [11,56], which indicates that they may be pairs of bona fide sister species, or alternatively they may represent genetic subdivisions within species generated for example by geographic or ecological isolation and genetic drift; (3) the *Rhizobium*-*Agrobacterium* clade is monophyletic in our topology, which would support the proposal of Young et al. [55] to include the species of the later genus in the former. However, taxon sampling for *Agrobacterium* is not comprehensive enough in our analysis to draw definitive conclusions on this controversial topic. Thorough phylogenetic and population genetic analyses based on DNA polymorphisms of more loci and large strain collections seem to be the most promising approaches to clarify these interesting issues of rhizobial systematics [46].

Acknowledgments

We greatly acknowledge the excellent technical support received from Margarita Sicardi de Mallorca (IVIC), Marco Antonio Rogel (CCG-UNAM), René Hernández (IBT-UNAM) and Scott Bringham (ASU, Tempe, AZ). The staff from the Botanical Garden of the Central University of Caracas is thanked for their help with the classification of the plant material. We warmly thank Kristina Lindström and Philippe de Lajudie for providing reference strains. J. Peter Young and Janet Sprent are acknowledged for their critical reading of the manuscript. Partial financial support was provided by UNAM-PAPIIT IN205802-3 (Mexico), INCO-DEV ICA-4-CT-2001-10057, Cicyt BMC2002-04126-C03-02 and Junta de Andalucía C03-049.

References

[1] E. Barrios, R. Herrera, Nitrogen cycling in a Venezuelan tropical seasonally flooded forest: soil nitrogen mineralization and nitrification, J. Trop. Ecol. 10 (1994) 399–416.

[2] P. de Lajudie, E. Laurent-Fulele, A. Willems, U. Tork, R. Coopman, M.D. Collins, K. Kersters, B. Dreyfus, M. Gillis, *Allorhizobium undicola* gen. nov., sp. nov., nitrogen fixing bacteria that efficiently nodulate *Neptunia natans* in Senegal, Int. J. Syst. Bacteriol. 48 (1998) 369–382.

[3] P. de Lajudie, A. Willems, G. Nick, F. Moreira, F. Molouba, B. Hoste, U. Torck, M. Neyra, M.D. Collins, K. Lindstrom, B. Dreyfus, M. Gillis, Characterization of tropical tree rhizobia and description of *Mesorhizobium plurifarium* sp. nov, Int. J. Syst. Bacteriol. 48 (1998) 369–382.

[4] L.R. Dice, Measures of the amount of ecological association between species, J. Ecol. 26 (1945) 297–302.

[5] F. Doignon-Bourcier, A. Willems, R. Coopman, G. Laguerre, M. Gillis, P. de Lajudie, Genotypic characterization of *Bradyrhizobium* strains nodulating small Senegalese legumes by 16S-23S rRNA intergenic gene spacers and amplified fragment length polymorphism fingerprint analyses, Appl. Environ. Microbiol. 66 (2000) 3987–3997.

[6] B. Dreyfus, J.L. Garcia, M. Gillis, Characterization of *Azorhizobium caulinodans* gen. nov., sp. nov., a stem-nodulationg nitroogen-fixing bacterium isolated from *Sesbania rostrata*, Int. J. Syst. Bacteriol. 38 (1988) 89–98.

[7] T. Eckhardt, A rapid method for the identification of plasmid desoxyribonucleic acid in bacteria, Plasmid 1 (1978) 584–588.

[8] P. Escobar-Páramo, A. Sabbagh, P. Darlu, O. Pradillon, C. Vaury, E. Denamur, G. Lecointre, Decreasing the effects of horizontal gene transfer on bacterial phylogeny: the *Escherichia coli* case study, Mol. Phylogenet. Evol. 30 (2004) 243–250.

[9] J. Felsenstein, Confidence limits on phylogenies: An approach using the bootstrap, Evolution 39 (1985) 783–791.

[10] J. Felsenstein, Inferring phylogenies, Sinauer Associates, INC, Sunderland, MA, 2004, p. 664.

[11] M.W. Gaunt, S.L. Turner, L. Rigottier-Gois, S.A. Lloyd-Macgilp, J.P. Young, Phylogenies of *atpD* and *recA* support the small subunit rRNA-based classification of rhizobia, Int. J. Syst. Evol. Microbiol. 51 (2001) 2037–2048.

[12] S. Guindon, O. Gascuel, A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood, Syst. Biol. 52 (2003) 696–704.
[13] K. Haukka, K. Lindström, J.P.W. Young, Three phylogenetic groups of *nodA* and *nifH* genes in *Sinorhizobium* and *Mesorhizobium* isolates from leguminous trees growing in Africa and Latin America, Appl. Environ. Microbiol. 64 (1998) 419–426.
[14] J.P. Huelsenbeck, F. Ronquist, R. Nielsen, J.P. Bollback, Bayesian inference of phylogeny and its impact on evolutionary biology, Science 294 (2001) 2310–2314.
[15] M.L. Izaguirre-Mayoral, O. Carballo, M. Sicardi de Mallorca, T. Oropeza, Quantitative analysis of the symbiotic N_2-fixation, non-structural carbohydrates and chlorophyll content in sixteen native legume species collected in different savanna sites, Symbiosis 12 (1992) 293–312.
[16] E.K. James, M.D. Loureiro, A. Pott, V.J. Pott, C.M. Martins, A.A. Franco, J.I. Sprent, Flooding-tolerant legume symbioses from the Brazilian Pantanal, New Phytol. 150 (2001) 723–738.
[17] E. Jaspers, J. Overmann, Ecological significance of microdiversity: identical 16S rRNA gene sequences can be found in bacteria with highly divergent genomes and ecophysiologies, Appl. Environ. Microbiol. 70 (2004) 4831–4839.
[18] P. Koponen, P. Nygren, A.M. Domenach, C. Le Roux, E. Saur, J.C. Roggy, Nodulation and dinitrogen fixation of legume trees in a tropical freshwater swamp forest in French Guiana, J. Trop. Ecol. 19 (2003) 655–666.
[19] G. Laguerre, S.M. Nour, V. Macheret, J. Sanjuan, P. Drouin, N. Amarger, Classification of rhizobia based on *nodC* and *nifH* gene analysis reveals a close phylogenetic relationship among *Phaseolus vulgaris* symbionts, Microbiology 147 (2001) 981–993.
[20] J. Lorquin, G. Lortet, M. Ferro, N. Mear, B. Dreyfus, J.C. Prome, C. Boivin, Nod factors from *Sinorhizobium saheli* and *S. teranga* bv. *sesbaniae* are both arabinosylated and fucosylated, a structural feature specific to *Sesbania rostrata* symbionts, Mol. Plant-Microbe Interact. 10 (1997) 879–890.
[21] B.L. Maidak, J.R. Cole, T.G. Lilburn, C.T. Parker Jr., P.R. Saxman, R.J. Farris, G.M. Garrity, G.J. Olsen, T.M. Schmidt, J.M. Tiedje, The RDP-II (Ribosomal Database Project), Nucleic Acids Res. 29 (2001) 173–174.
[22] D.P. Martin, C. Williamson, D. Posada, RDP2: recombination detection and analysis from sequence alignments Bioinformatics 21 (2005) 260–262.
[23] R. Nichols, Gene trees and species trees are not the same Trends, Ecol. Evol. 16 (2001) 358–364.
[24] J.A. Nylander, F. Ronquist, J.P. Huelsenbeck, J.L. Nieves-Aldrey, Bayesian phylogenetic analysis of combined data, Syst. Biol. 53 (2004) 47–67.
[25] D. Posada, T.R. Buckley, Model selection and model averaging in phylogenetics: advantages of akaike information criterion and bayesian approaches over likelihood ratio tests, Syst. Biol. 53 (2004) 793–808.
[26] D. Posada, K.A. Crandall, MODELTEST: testing the model of DNA substitution, Bioinformatics 14 (1998) 817–818.
[27] D. Posada, K.A. Crandall, Evaluation of methods for detecting recombination from DNA sequences: Computer simulations, Proc. Natl. Acad. Sci. USA 98 (2001) 13757–13762.
[28] J. Qian, M.A. Parker, Contrasting *nifD* and ribosomal gene relationships among *Mesorhizobium* from *Lotus oroboides* in northern Mexico, Syst. Appl. Microbiol. 25 (2002) 68–73.
[29] J.L.W. Rademaker, H.J.M. Aarts, P. Vinuesa, Molecular typing of environmental isolates, In: A.M. Osborn (Ed.), Molecular Microbial Ecology, BIOS Scientific Publisher, Oxford, UK, 2005, pp. 1–20 pp. 97–134.
[30] F. Ronquist, J.P. Huelsenbeck, MrBAYES 3: Bayesian phylogenetic inference under mixed models Bioinformatics 19 (2003) 1572–1574.
[31] N. Saitou, M. Nei, The neighbor-joining method: A new method for reconstructing phylogenetic trees, Mol. Biol. Evol. 4 (1987) 406–425.
[32] M.O. Salminen, J.K. Carr, D.S. Burke, F.E. McCutcham, Identification of breakpoints in intergenotypic recombinants of HIV type 1 by Bootscanning, AIDS Res. Hum. Retroviruses 11 (1995) 1423–1425.
[33] H. Sawada, L.D. Kuykendall, J.M. Young, Changing concepts in the systematics of bacterial nitrogen-fixing legume symbionts, J. Gen. Appl. Microbiol. 49 (2003) 155–179.
[34] S. Sawyer, Statistical tests for detecting gene conversion, Mol. Biol. Evol. 6 (1998) 526–538.
[35] M.H. Schierup, J. Hein, Consequences of recombination on traditional phylogenetic analysis, Genetics 156 (2000) 879–891.
[36] H. Shimodaira, M. Hasegawa, Multiple comparisons of log-likelihoods with applications to phylogenetic inference, Mol. Biol. Evol. 16 (1999) 1114–1116.
[37] P. Somasegaran, H.J. Hoben, Handbook for rhizobia: Methods in legume-*Rhizobium* technology, Springer, Heidelberg, Germany, 1994.
[38] Y. Suzuki, G.V. Glazko, M. Nei, Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics, Proc. Natl. Acad. Sci. USA 99 (2002) 16138–16143.
[39] D.L. Swofford, PAUP*: Phylogenetic Analysis Using Parsimony and Other Methods (software), Sinuauer Associates, Sunderland, MA, 2002.
[40] Z. Tan, T. Hurek, P. Vinuesa, P. Muller, J.K. Ladha, B. Reinhold-Hurek, Specific detection of *Bradyrhizobium* and *Rhizobium* srains colonizing rice (*Oryza sativa*) roots by 16S-23S ribosomal DNA intergenic spacer- targeted PCR, Appl. Environ. Microbiol. 67 (2001) 3655–3664.
[41] Z. Terefework, G. Nick, S. Suomalainen, L. Paulin, K. Lindstrom, Phylogeny of *Rhizobium galegae* with respect to other rhizobia and agrobacteria, Int. J. Syst. Bacteriol. 48 (Pt 2) (1998) 349–356.
[42] P. van Berkum, Z. Terefework, L. Paulin, S. Suomalainen, K. Lindstrom, B.D. Eardly, Discordant phylogenies within the *rrn* loci of rhizobia, J. Bacteriol. 185 (2003) 2988–2998.
[43] P. Vinuesa, M. León-Barrios, C. Silva, A. Willems, A. Jarabo-Lorenzo, R. Pérez-Galdona, D. Werner, E. Martínez-Romero, *Bradyrhizobium canariense* sp.

nov., an acid-tolerant endosymbiont that nodulates endemic genistoid legumes (Papilionoideae:Genisteae) growing in the Canary Islands, along with *B. japonicum* bv. genistearum, *Bradyrhizobium* genospecies α and *Bradyrhizobium* genospecies β, Int. J. Syst. Evol. Microbiol. 55 (2005) 569–575.

[44] P. Vinuesa, J.L.W. Rademaker, F.J. de Bruijn, D. Werner, Genotypic characterization of *Bradyrhizobium* strains nodulating endemic woody legumes of the Canary Islands by PCR-restriction fragment length polymorphism analysis of genes encoding 16S rRNA (16S rDNA) and 16S-23S rDNA intergenic spacers, repetitive extragenic palindromic PCR genomic fingerprinting and partial 16S rDNA sequencing, Appl. Environ. Microbiol. 64 (1998) 2096–2104.

[45] P. Vinuesa, C. Silva, Species delineation and biogeography of symbiotic bacteria associated with cultivated and wild legumes, In: D. Werner (Ed.), Biological Resources and Migration, Springer, Berlin, 2004, pp. 143–161.

[46] P. Vinuesa, C. Silva, D. Werner, E. Martínez-Romero, Population genetics and phylogenetic inference in bacterial molecular systematics: the roles of migration and recombination in *Bradyrhizobium* species cohesion and delineation, Mol. Phylogenet. Evol. 34 (2005) 29–54.

[47] E.T. Wang, F.L. Kan, Z.Y. Tan, I. Toledo, W.X. Chen, E. Martinez-Romero, Diverse *Mesorhizobium plurifarium* populations native to Mexican soils, Arch. Microbiol. 180 (2003) 444–454.

[48] E.T. Wang, E. Martinez-Romero, *Sesbania herbacea-Rhizobium huautlense* nodulation in flooded soils and comparative characterization of *S. herbacea*-nodulating rhizobia in different environments, Microb. Ecol. 40 (2000) 25–32.

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

[50] Y. Wang, Z. Zhang, Comparative sequence analyses reveal frequent occurrence of short segments containing an abnormally high number of non-random base variations in bacterial rRNA genes, Microbiology 146 (2000) 2845–2854.

[51] B.S. Weir, S.J. Turner, W.B. Silvester, D.C. Park, J.M. Young, Unexpectedly diverse *Mesorhizobium* strains and *Rhizobium leguminosarum* nodulate native legume genera of New Zealand, while introduced legume weeds are nodulated by *Bradyrhizobium* species, Appl. Environ. Microbiol. 70 (2004) 5980–5987.

[52] W.G. Weisburg, S.M. Barns, D.A. Pelletie, D.J. Lane, 16S ribosomal DNA amplification for phylogenetic study, J. Bacteriol. 173 (1991) 697–703.

[53] X. Xia, Z. Xie, DAMBE: Software package for data analysis in molecular biology and evolution, J. Hered. 92 (2001) 371–373.

[54] Z. Yang, Among-site rate variation and its impact on phylogenetic analyses, Trends Ecol. Evol. 11 (1996) 367–372.

[55] J.M. Young, L.D. Kuykendall, E. Martinez-Romero, A. Kerr, H. Sawada, A revision of *Rhizobium* Frank 1889, with an emended description of the genus, and the inclusion of all species of *Agrobacterium* Conn 1942 and *Allorhizobium undicola* de Lajudie et al. 1998 as new combinations: *Rhizobium radiobacter*, *R. rhizogenes*, *R. rubi*, *R. undicola* and *R. vitis*, Int. J. Syst. Evol. Microbiol. 51 (2001) 89–103.

[56] J.M. Young, D.C. Park, B.S. Weir, Diversity of 16S rDNA sequences of *Rhizobium* spp. implications for species determinations, FEMS Microbiol. Lett. 238 (2004) 125–131.

[57] J.L. Zurdo-Pineiro, E. Velazquez, M.J. Lorite, G. Brelles-Marino, E.C. Schroder, E.J. Bedmar, P.F. Mateos, E. Martinez-Molina, Identification of fast-growing rhizobia nodulating tropical legumes from Puerto Rico as *Rhizobium gallicum* and *Rhizobium tropici*, Syst. Appl. Microbiol. 27 (2004) 469–477.