



HAL
open science

Conflicting phylogeographic patterns in rRNA and nifD indicate regionally restricted gene transfer in *Bradyrhizobium*.

Matthew A. Parker, Benedicte Lafay, Jeremy J. Burdon, Peter van Berkum

► To cite this version:

Matthew A. Parker, Benedicte Lafay, Jeremy J. Burdon, Peter van Berkum. Conflicting phylogeographic patterns in rRNA and nifD indicate regionally restricted gene transfer in *Bradyrhizobium*. Microbiology, 2002, 148 (Pt 8), pp.2557-65. hal-00412903

HAL Id: hal-00412903

<https://hal.science/hal-00412903v1>

Submitted on 3 Jun 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Conflicting phylogeographic patterns in rRNA and *nifD* indicate regionally restricted gene transfer in *Bradyrhizobium*

Matthew A. Parker,¹ Benedicte Lafay,^{2†} Jeremy J. Burdon² and Peter van Berkum³

Author for correspondence: Matthew A. Parker. Tel: +1 607 777 6283. Fax: +1 607 777 6521.
e-mail: mparker@binghamton.edu

¹ Department of Biological Sciences, State University of New York, Binghamton, NY 13902, USA

² Centre for Plant Biodiversity Research, CSIRO Plant Industry, Canberra ACT 2601, Australia

³ Soybean and Alfalfa Research Laboratory, USDA, ARS, HH-4, Bldg010, BARC-West, 10300 Baltimore Blvd, Beltsville, MD 20705, USA

Major differences in evolutionary relationships of the 16S rRNA gene and the nitrogenase α -subunit gene (*nifD*) were observed among 38 strains of *Bradyrhizobium* sp. nodule bacteria from North America, Central America, Asia and Australia. Two lineages were evident in the 16S rRNA phylogeny representing strains related to *Bradyrhizobium japonicum* (29 isolates) or *Bradyrhizobium elkanii* (9 isolates). Both clades were distributed across most or all of the geographic regions sampled. By contrast, in the *nifD* tree almost all isolates were placed into one of three groups each exclusively composed of taxa from a single geographic region (North Temperate, Central America or Australia). Isolates that were closely related or identical in gene sequence at one locus often had divergent sequences at the other locus and a partition homogeneity test indicated that the 16S rRNA and *nifD* phylogenies were significantly incongruent. No evidence for any gene duplication of *nifD* was found by Southern hybridization analysis on a subset of the strains, so unrecognized paralogy is not likely to be responsible for the discrepancy between 16S rRNA and *nifD* tree topologies. These results are consistent with a model whereby geographic areas were initially colonized by several diverse 16S rRNA lineages, with subsequent horizontal gene transfer of *nifD* leading to increased *nifD* sequence homogeneity within each regional population.

Keywords: geographic variation, nitrogenase, nodule bacteria, *Rhizobiaceae*

INTRODUCTION

The prevalence of gene transfer in natural populations of root nodule bacteria remains highly uncertain. There is evidence for plasmid exchange in populations of *Rhizobium leguminosarum* associated with cultivated peas, because different chromosomal lineages may carry highly similar symbiotic plasmids (Young & Wexler, 1988; Rigottier-Gois *et al.*, 1998). Sullivan *et al.* (1995, 1996) also demonstrated that chromosomal symbiotic genes were transferred from an inoculant strain of *Mesorhizobium loti* to non-symbiotic bacteria, enabling

them to nodulate a newly established pasture population of *Lotus corniculatus*. However, in agriculture the opportunity for gene transfer could be much higher than in natural plant communities, where rhizobia may occur at lower numbers and may be separated by microhabitat or by differential nodulation compatibility with different legume hosts (Wernegreen *et al.*, 1997; Wernegreen & Riley, 1999). The uncertainty of gene exchange among natural populations of rhizobia was highlighted by the failure to find evidence for symbiotic plasmid transfer in *R. leguminosarum* associated with wild clovers in California (Wernegreen *et al.*, 1997).

Incongruence between phylogenetic trees for different loci also provides evidence for horizontal gene transfer (Dykhuisen & Green, 1991; Zhou *et al.*, 1997). Incongruent gene trees have been reported within the genera *Rhizobium*, *Sinorhizobium* and *Mesorhizobium* (Haukka *et al.*, 1998; Wernegreen & Riley, 1999; Turner & Young, 2000; Laguerre *et al.*, 2001; Qian & Parker, 2002). However, it is important to investigate additional

† Present address: Centre d'Etudes sur le Polymorphisme des Micro-Organismes, UMR CNRS-IRD 9926, 911 Avenue Agropolis – BP 64501, 34394 Montpellier Cedex 5, France.

Abbreviations: ML, maximum-likelihood; MP, maximum-parsimony; NJ, neighbour-joining.

The GenBank accession numbers for the *nifD* sequences determined in this work are AF484254–AF484287.

examples from a variety of geographic regions, to better characterize the prevalence of gene transfer in natural environments.

Bacteria in the genus *Bradyrhizobium* are important symbionts for diverse legume taxa throughout the world, but no studies have been done with a broad geographic sample of isolates to compare relationships for sequences of 16S rRNA and other gene loci. In this study, we sequenced 822 bp of the gene for the α -subunit of nitrogenase (*nifD*) in isolates of *Bradyrhizobium* from Asia, North America, Central America and Australia that had previously been characterized for 16S rRNA relationships (Lafay & Burdon, 1998, 2001; Parker, 1999, 2000, 2001; Parker & Lunk, 2000; van Berkum & Fuhrmann, 2000). Our objectives were (1) to evaluate the degree of phylogenetic tree congruence across the two loci and (2) to compare the geographic structure of bacterial relationships revealed by 16S rRNA vs *nifD* sequence variation. Our results indicate very different phylogeographic patterns for these two loci, and suggest that differences in gene transfer have greatly altered their relative genealogical history.

METHODS

Bacterial strains. *nifD* genes were sequenced for 35 *Bradyrhizobium* strains previously characterized for 16S rRNA variation (Table 1). For three other strains (*Bradyrhizobium japonicum* USDA 110, jwc91-2, Tv2a-2), both *nifD* and 16S rRNA data were already available from GenBank (Z35330, X01045, AF178437, AF216780, AF285776, AF285777). The strains analysed included (1) the type strains of the three named species in the genus (*B. japonicum*, *Bradyrhizobium liaoningense*, *Bradyrhizobium elkanii*) and 14 other strains of these taxa from the USDA culture collection (van Berkum & Fuhrmann, 2000), (2) 11 strains of *Bradyrhizobium* isolated from woody legumes endemic to south-eastern Australia (Lafay & Burdon, 1998, 2001), (3) three isolates from the eastern North American legumes *Amphicarpaea* and *Apios* (Parker, 1999) and (4) seven isolates from six genera of legumes growing on Barro Colorado Island, Panama (Parker & Lunk, 2000; Parker, 2000, 2001).

DNA amplification and sequencing. DNA was purified from bacterial cells by the protocol of Lafay & Burdon (1998) or by the use of a Qiagen DNA extraction kit (van Berkum & Fuhrmann, 2000). Approximately 830 bp of *nifD* was amplified using primers nifp11 and nifp12 (Parker, 2000) which yield a fragment that starts at codon 28 of the *B. japonicum* USDA 110 *nifD* gene (GenBank X01045). For certain isolates that failed to amplify strongly with these primers, alternative primers homologous to nearby portions of *nifD* were used instead (a complete list of both amplification and internal sequencing primers is available by request from M. Parker). PCR used 25 μ l reaction mixtures containing 10 mM Tris buffer with 0.1% Triton X-100, 50 mM KCl, 1.5 mM MgCl₂, 0.2 mM each dNTP, 0.5 μ M each primer, 0.5 μ l genomic DNA and 0.5 U *Taq* polymerase. Tubes were incubated for 90 s at 94 °C and then subjected to 33 cycles of 94 °C (20 s), 56 °C (20 s) and 72 °C (80 s), with a final extension of 4 min at 72 °C. PCR products were sequenced on both strands using Applied Biosystems model 310 and 377 automated sequencers with protocols recommended by the manufacturer.

Phylogenetic analyses. Sequences were aligned using CLUSTAL W (Thompson *et al.*, 1994) and trees were

constructed by maximum-parsimony (MP), neighbour-joining (NJ) and maximum-likelihood (ML) methods using the PAUP software (version 4.0b1, from D. L. Swofford, Smithsonian Institution, Washington, DC, USA). To determine the degree of statistical support for branches in the phylogeny (Felsenstein, 1985), 1000 bootstrap replicates of the data were analysed (100 only for ML). NJ analysis (Saitou & Nei, 1987) used HKY85 distances (Hasegawa *et al.*, 1985). Two genera of α -Proteobacteria (*Azospirillum brasilense*, *Rhodobacter capsulatus*), for which both 16S rRNA and *nifD* sequences were available in GenBank (Z29617, D16428, M64344, M15270), served as outgroups in all analyses.

RESULTS

16S rRNA relationships

The 38 strains of *Bradyrhizobium* were placed into two groups by parsimony analysis of aligned 16S rRNA sequences (Fig. 1). One group consisted of 29 isolates related to *B. japonicum* and the other was comprised of 9 strains related to *B. elkanii*. These same two lineages have been identified in all previous analyses of 16S rRNA sequences regardless of the method of phylogenetic analysis (e.g. Young & Haukka, 1996; Lafay & Burdon, 1998; van Berkum & Fuhrmann, 2000; Parker & Lunk, 2000; Parker, 2001; Willems *et al.*, 2001). Identical groups were identified in NJ and ML analyses of 16S rRNA data for these strains. Bootstrap support values for the *B. japonicum* and *B. elkanii* lineages were 99 and 100%, respectively, for NJ trees, and 72 and 60% in the ML analysis.

There was insufficient sequence variation in the 16S rRNA gene to resolve relationships within the two main groups. For example, the median number of pairwise nucleotide differences among 29 isolates in the *B. japonicum* group was only 11 (<0.8% of the 1410 bp 16S rRNA region analysed), producing a large basal multifurcation on the bootstrap consensus tree (Fig. 1). Four sets of isolates had identical 16S rRNA sequences (apart from 1–2 bp length differences in a few isolates, which were excluded from the phylogenetic analyses). In the *B. japonicum* group, USDA 62, 122 and 129 were identical, as were USDA 123 and 127, and *B. liaoningense* and *B. japonicum* USDA 135. Within the *B. elkanii* group, USDA 31, 94, 130 and isolate Cj3-3 also had identical 16S rRNA sequences, even though Cj3-3 was from Panama and the USDA strains originated from North America. Several other cases were observed where strains with closely similar 16S rRNA sequences were obtained from separate regions. For example, Australian strain 5329H had only one substitution relative to *B. japonicum* USDA 38 from North America, and Australian strain 5111P showed 2 nt differences from North American strain th-b2.

Seven of the ten Australian isolates, and four of the six Panamanian isolates, each formed an apparent group within the *B. japonicum* clade. Apart from these small groups, there was little correspondence between geographic origin of the strains and their phylogeny. For example, isolates from North America, Central America and Australia were represented in both of the major

Table 1. Bacterial strains

Strain	Host	Origin	Reference or source
<i>B. japonicum</i>			
USDA 4	<i>Glycine max</i>	USA	USDA collection
USDA 6 ^T	<i>Glycine max</i>	Japan	USDA collection
USDA 38	<i>Glycine max</i>	Japan	USDA collection
USDA 62	<i>Glycine max</i>	USA	USDA collection
USDA 110	<i>Glycine max</i>	USA	USDA collection
USDA 122	<i>Glycine max</i>	USA	USDA collection
USDA 123	<i>Glycine max</i>	USA	USDA collection
USDA 124	<i>Glycine max</i>	USA	USDA collection
USDA 127	<i>Glycine max</i>	USA	USDA collection
USDA 129	<i>Glycine max</i>	USA	USDA collection
USDA 135	<i>Glycine max</i>	USA	USDA collection
<i>B. liaoningense</i>			
USDA 3622 ^T	<i>Glycine max</i>	China	Xu <i>et al.</i> (1995)
<i>B. elkanii</i>			
USDA 31	<i>Glycine max</i>	USA	USDA collection
USDA 46	<i>Glycine max</i>	USA	USDA collection
USDA 76 ^T	<i>Glycine max</i>	USA	USDA collection
USDA 94	<i>Glycine max</i>	USA	USDA collection
USDA 130	<i>Glycine max</i>	USA	USDA collection
<i>Bradyrhizobium</i> sp.			
jwc91-2	<i>Amphicarpaea bracteata</i>	USA	Parker (1999)
th-b2	<i>Amphicarpaea bracteata</i>	USA	Parker (1999)
ApB16	<i>Apios americana</i>	USA	Parker (1999)
Mm1-3	<i>Machaerium milleflorum</i>	Panama	Parker & Lunk (2000)
Pe4	<i>Platypodium elegans</i>	Panama	Parker & Lunk (2000)
Pe1-3	<i>Platypodium elegans</i>	Panama	Parker & Lunk (2000)
Tv2a-2	<i>Tachigali versicolor</i>	Panama	Parker (2000)
Ec3-3	<i>Erythrina costaricensis</i>	Panama	Parker (2001)
Da3-1	<i>Desmodium axillare</i>	Panama	Parker (2001)
Cj3-3	<i>Clitoria javitensis</i>	Panama	Parker (2001)
5028A	<i>Bossiaea ensata</i>	Australia	Lafay & Burdon (1998)
5057B	<i>Bossiaea foliosa</i>	Australia	Lafay & Burdon (1998)
5792C	<i>Oxylobium ellipticum</i>	Australia	Lafay & Burdon (1998)
5563D	<i>Podolobium ilicifolium</i>	Australia	Lafay & Burdon (1998)
5029F	<i>Bossiaea ensata</i>	Australia	Lafay & Burdon (1998)
5680G	<i>Bossiaea ensata</i>	Australia	Lafay & Burdon (1998)
5329H	<i>Goodia lotifolia</i>	Australia	Lafay & Burdon (1998)
5915J	<i>Bossiaea foliosa</i>	Australia	Lafay & Burdon (1998)
5493M	<i>Mirbelia rubiifolia</i>	Australia	Lafay & Burdon (1998)
1808N	<i>Acacia dealbata</i>	Australia	Lafay & Burdon (2001)
5111P	<i>Daviesia leptophylla</i>	Australia	Lafay & Burdon (1998)

groups. To analyse the extent to which geographic origin was distributed non-randomly on the tree of Fig. 1, the location of terminal taxa was randomly permuted 1000 times while holding tree topology constant (Maddison & Slatkin, 1991). The minimum number of migration events consistent with the phylogeny was then inferred by MP criteria (Slatkin & Maddison, 1989). This provides a null distribution for the number of migration events that can be compared to the pattern in the actual tree. This analysis indicated that chance alone would generate as much geographic clustering as ob-

served (Fig. 1) in about 375 out of 1000 trials. Thus, the null hypothesis that phylogenetically related isolates are distributed at random across regions cannot be rejected for 16S rRNA sequences.

***nifD* relationships**

PCR using primers for *nifD* yielded an 822 bp amplification product with the exception of 10 of the 11 Australian isolates, which generated a 3 bp shorter fragment. All Australian isolates except strain 5111P

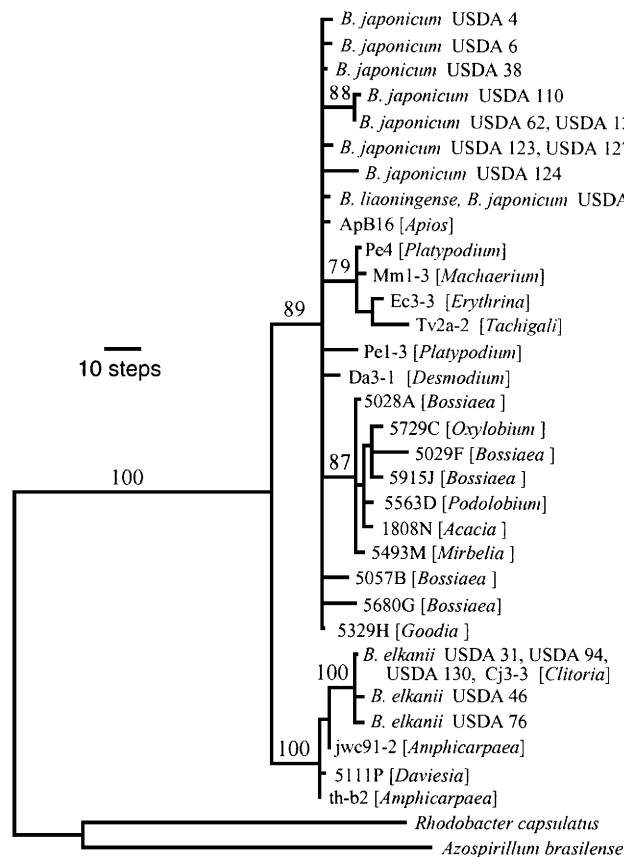


Fig. 1. MP bootstrap majority-rule consensus tree for 16S rRNA sequences from 38 *Bradyrhizobium* isolates. Numbers adjacent to branch points are bootstrap percentages ($n = 1000$ replicates).

had a deletion corresponding to codon 220 of the *B. japonicum* USDA 110 *nifD* gene. This insertion/deletion polymorphism was excluded from the data for phylogenetic analyses, but it does provide additional support for a close genealogical relationship of *nifD* genes among the Australian bradyrhizobia (see below). The fraction of polymorphic sites in first, second and third codon positions was 0.157, 0.077 and 0.901, respectively. Mutational saturation at third codon positions can cause misleading results in parsimony analysis of highly divergent sequences (Cunningham, 1997). However, the transition/transversion ratio for third position substitutions (2.76) was not depressed relative to first and second positions (1.72, 1.22), indicating that the phylogenetic signal associated with third position polymorphisms had not been erased by multiple substitutions. To investigate this issue, MP analyses were done on both the full dataset and on a reduced dataset where third codon positions with transitions were excluded.

MP analysis of the full *nifD* dataset yielded 84 trees that differed primarily in the placement of closely related strains within terminal clades. In the consensus tree, nearly all *nifD* sequences clustered into groups exclusively composed of taxa from a single geographic region (Fig. 2). For example, five Panamanian isolates from diverse legume hosts formed a group in 100% of bootstrap replicates. Similarly, all of the Australian isolates formed a highly supported group. All *B. elkanii* strains had identical *nifD* sequences, and were placed with all of the other North American strains (jwc91-2,

th-b2, ApB16 and 10 *B. japonicum* strains) in 97% of bootstrap replicates. The three Asian soybean strains (USDA 6, USDA 38, *B. liaoningense*) each had *nifD* sequences that were identical to those of certain North American *B. japonicum* strains (e.g. USDA 4 and USDA 135). Therefore, the entire set of *Bradyrhizobium* strains originating from North America and Asia was labelled 'North Temperate' (Fig. 2). An MP analysis excluding 121 third codon sites with transitions yielded an identical tree topology as in Fig. 2, with bootstrap values of 99, 89 and 100% for the groups formed by strains originating from North Temperate, Australia and Panama, respectively. The same three highly supported groups and precisely the same branching order within groups as in Fig. 2 were observed in NJ and ML analyses. Bootstrap values for the North Temperate, Australia and Panama groups were all 100% for the NJ tree and were 99, 98 and 100%, respectively, for the ML analysis.

The *nifD* sequences of Tv2a-2 and Cj3-3 from Panama were divergent from all other sequences obtained for strains from that region. Their inferred phylogenetic positions were dependent on the method of tree construction. In a NJ analysis, Cj3-3 was placed as a basal member of the North Temperate group (with 75% bootstrap support), while Tv2a-2 was placed in an unresolved trichotomy with the North Temperate and Australia groups. Both these isolates were placed in an unresolved multifurcation with the North Temperate and Australian groups by ML analysis. Constraining these two strains to form a group with the remaining five

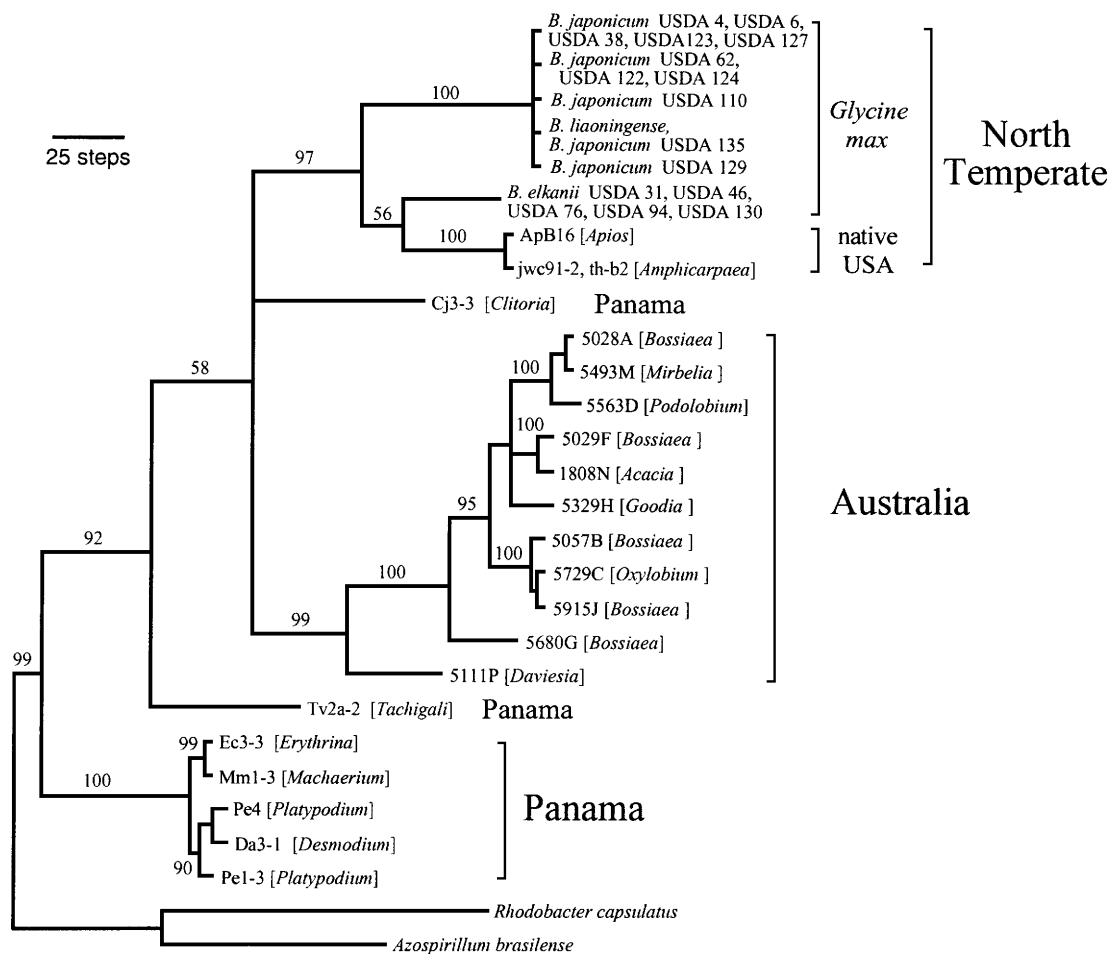


Fig. 2. MP bootstrap majority-rule consensus tree for *nifD* sequences from 38 *Bradyrhizobium* isolates. Numbers adjacent to branch points are bootstrap percentages ($n = 1000$ replicates).

Panama strains produced MP trees that were significantly less likely than the phylogenies reconstructed without this topological constraint according to the Shimodaira–Hasegawa test (Shimodaira & Hasegawa, 1999). This indicates that both these strains possess *nifD* sequences that diverge significantly from the remaining strains originating from Panama.

In summary, 36 of the 38 *Bradyrhizobium nifD* sequences clustered into three geographic groups that were highly supported by all methods of phylogenetic analysis, and the relationships of two others (from isolates Tv2a-2 and Cj3-3) must be considered unresolved. The extent of geographic clustering was analysed by a permutation test (Maddison & Slatkin, 1991, as described above). In the actual tree, the geographic distribution of all *nifD* sequences requires a hypothesis of just four changes of location throughout the entire phylogeny. However, in 1000 randomly permuted datasets, the number of inferred migration events ranged from 8 to 16 (median 13). Thus, phylogenetically related *nifD* sequences showed highly significant geographic clustering; there is less than one chance in a thousand of obtaining the pattern of regional clades (Fig. 2) under

the null hypothesis that geographic origin and genealogy are independent.

Are 16S rRNA and *nifD* trees congruent?

Comparison of 16S rRNA and *nifD* relationships indicated that there were several major discrepancies. The primary division between *B. japonicum* and *B. elkanii* lineages in the 16S rRNA data (Fig. 1) was missing for *nifD* (Fig. 2). For example, strain Cj3-3 had a 16S rRNA sequence identical to that of *B. elkanii* USDA 94, but had a *nifD* sequence that was divergent from all *B. elkanii* strains. The 16S rRNA gene of Australian strain 5111P differed by only two to three substitutions from certain North American isolates in the *B. elkanii* group (Fig. 1), but the *nifD* gene of strain 5111P was similar to those of other Australian strains (Fig. 2) related to *B. japonicum* in the 16S rRNA analysis. Isolate ApB16 was a close relative of several isolates in the 16S rRNA *B. japonicum* group (e.g. it differed from USDA 38 and 5329H by one and two substitutions, respectively). However, the *nifD* sequence of ApB16 was highly similar to those of two North American isolates (jwc91-2, th-b2) that were placed with

B. elkanii in the 16S rRNA analysis. The partition homogeneity test of Farris *et al.* (1995) indicated that the MP trees for 16S rRNA and *nifD* were significantly incongruent ($P < 0.001$). Thus, the genealogical histories of these two genes are demonstrably different for this set of *Bradyrhizobium* strains.

Gene duplication could potentially lead to an incorrect inference regarding phylogenetic incongruence across loci. For example, if multiple copies of either 16S rRNA or *nifD* genes are present, then a mixed sample that included non-orthologous gene sequences could misrepresent the true pattern of relationships (Doyle & Davis, 1998; Wendel & Doyle, 1998). However, all existing evidence suggests that *Bradyrhizobium* possesses a single rRNA gene region (Kundig *et al.*, 1995; van Berkum *et al.*, 1998; Klappenbach *et al.*, 2000). In strain USDA 110 of *B. japonicum*, *nifD* also appears to be a single-copy gene (Kundig *et al.*, 1993; Gottfert *et al.*, 2001), but information is very limited for other strains. Therefore, Southern hybridization analysis was done with ten of the *Bradyrhizobium* strains using HindIII and EcoRI digests of genomic DNA and 822 bp of the *B. japonicum* USDA 110 *nifD* gene as probe. For HindIII digests, a single hybridization signal was observed in all isolates, whereas in EcoRI digests six of the ten isolates had a single band and the remaining four isolates (jwc91-2, th-b2, ApB16 and Cj3-3) displayed two hybridization signals. Inspection of *nifD* sequences for these four strains indicated that they had an EcoRI site in the middle of the gene, thus accounting for the presence of two genomic restriction fragments with *nifD* sequence homology. These results imply that all strains most likely have a single copy of *nifD*. Thus, the incongruence of tree topology in Figs 1 and 2 is unlikely to be an artefact caused by the inclusion of paralogous *nifD* genes in the analysis.

DISCUSSION

Our main finding was that ribosomal and *nifD* genes had strongly discordant phylogeographic patterns. Strains with similar or identical 16S rRNA sequences commonly occurred in widely separated regions and there was no significant geographic structure in the MP tree (Fig. 1). This implies that there has been extensive intercontinental migration over the time scale of evolutionary divergence at this locus. By contrast, strains with similar or identical *nifD* sequences were almost never shared across regions, and related strains showed a significant tendency to group according to geographic origin (Fig. 2). This geographic pattern implies that regional populations have been effectively isolated during the evolution of sequence divergence at the *nifD* locus.

The only cases where identical *nifD* sequences were found among strains originating from different regions involved symbionts of soybean (*Glycine max*). USDA 6 and USDA 38, both from Japan, had *nifD* sequences that were identical to several North American *B. japonicum* strains (USDA 4, USDA 123, USDA 127). Also, the *nifD*

sequence of *B. liaoningense* from China was identical to *B. japonicum* USDA 135 from Iowa. These possibly represent cases of recent human-mediated dispersal, since there are historical records documenting the deliberate introduction of bacterial symbionts from Asia during the early years of soybean cultivation in North America (e.g. Wilson, 1934). Apart from these two cases, the complete absence of any closely related *nifD* sequences from different regions was noteworthy. This suggests that inter-regional migration of *Bradyrhizobium* has been negligible over recent evolutionary time.

Host specificity could be a factor contributing to the strong regional differentiation of *nifD* sequences. If bradyrhizobial strains tended to be narrowly specialized for nodulation of specific legume taxa, then the restricted geographic range shown by most legume hosts would limit the distribution of bacterial lineages. For example, the hosts for three Panama isolates were legumes in the tribe *Dalbergieae* (*Machaerium*, *Platypodium*), which is largely restricted to tropical America, and all but one of the Australian isolates came from legume genera in the tribes *Mirbelieae* and *Bossiaeeae* which are endemic to Australia and nearby south-east Asia. Existing data on host specificity are too limited to fully analyse this problem. All of the Australian and Panamanian strains have the ability to form nodules on *Macroptilium atropurpureum* (tribe *Phaseoleae*; Parker & Lunk, 2000; Parker, 2000, 2001; B. Lafay, unpublished data), which indicates that they are not strictly specialized on their original legume host. Among the Panama isolates, apparently identical bacterial genotypes are shared by legume hosts from three separate legume tribes (Parker & Lunk, 2000; Parker, 2001), again suggesting that these strains are not highly host specific. An important direction for future research will be to analyse bacteria sampled from related legumes in different regions, so that host taxonomic group is not confounded with geographic origin.

The differential rate of evolution of 16S rRNA and *nifD* may be partially responsible for variation in phylogeographic patterns. The 16S rRNA gene evolves slowly (Woese, 1987), with an estimated substitution rate per site of 0.012–0.018 per 100 million years in non-endosymbiotic bacteria (Clark *et al.*, 1999). Assuming that substitutions occur according to a Poisson process (Werren *et al.*, 1995), the upper 95% confidence limit for divergence time in cases where strains showed no nucleotide differences throughout the entire 1410 bp 16S rRNA fragment would range from 5.9 to 8.9 million years. Thus, pairs of isolates in different regions that descended from a common ancestor more than 5 million years in the past would often retain identical 16S rRNA sequences. No estimates of evolutionary rate for *nifD* have been made. However, a substantial difference in evolutionary rate is implied since *nifD* had more than a sixfold higher frequency of polymorphic sites than 16S rRNA (0.376 vs 0.059). If a sixfold higher evolutionary rate is assumed for *nifD*, then there is less than a 5% chance that pairs of isolates separated for more than 1.5 million years would retain identical sequences. A higher

evolutionary rate for *nifD* would contribute to the evolution of distinctive regional clades in geographically isolated populations (Fig. 2), despite a lack of strong phylogeographic structure in the more slowly evolving 16S rRNA gene (Fig. 1).

Nevertheless, differences in evolutionary rate alone cannot account for the significantly different tree topologies for the two genes. Figs 1 and 2 reveal numerous examples where strains that were closely related for one gene had very divergent sequences in the other, and the overall structure of strain relationships was significantly discordant across loci. For example, the most basal split in the 16S rRNA tree involved divergence into two lineages related to *B. japonicum* and *B. elkanii*, which were both represented in most of the regions sampled (Fig. 1). By contrast, in the *nifD* tree there was no indication of a basal split into two lineages each shared by multiple geographic areas (Fig. 2).

Horizontal gene transfer at one or both loci is the most likely process responsible for these conflicting phylogenies (Spratt & Maiden, 1999). For bacterial ribosomal genes, several cases have been reported for transfer events involving entire genes (Yap *et al.*, 1999) or portions of genes (Eardly *et al.*, 1996; Wang & Zhang, 2000; Parker, 2001). However, our data suggest that horizontal transfers of *nifD* may have been more prevalent. Isolates from the same geographic region almost invariably resembled one another for *nifD* sequences, even when they were not close relatives at the 16S rRNA locus. This would suggest a scenario where each region was initially colonized by multiple *Bradyrhizobium* lineages with divergent 16S rRNA sequences, followed by a process of lateral *nifD* gene transfer. More frequent transfer of *nifD* than 16S rRNA would account for the within-region homogeneity of *nifD* sequences (Fig. 2) across lineages that have retained divergent 16S rRNA genes (Fig. 1).

For symbiotic loci such as the common *nod* genes, it has been suggested that gene trees match the phylogeny of host legumes and often show discrepancies from bacterial 16S rRNA relationships (Ueda *et al.*, 1995; Wernegreen & Riley, 1999; Laguerre *et al.*, 2001). The extent to which *nifD* may be expected to match host plant phylogeny is uncertain. Differences commonly exist between *nod* and *nif* gene trees (Haukka *et al.*, 1998; Laguerre *et al.*, 2001) and our *nifD* tree showed no clear correspondence to the phylogenetic relationships of host legumes. For example, the legume *Tachigali* belongs to a lineage in the subfamily *Caesalpinioideae* that gave rise to the subfamily *Mimosoideae* (Doyle *et al.*, 1997). Yet the sole bradyrhizobial isolate from a mimosoid legume (strain 1808N from *Acacia*) did not cluster with the *Tachigali* isolate (Tv2a-2), but rather, grouped with isolates from distantly related legumes (Fig. 2). Also, the host for strain Ec3-3 from Panama is in the same subtribe as that for isolate ApB16 from North America (subtribe *Erythrininae*, tribe *Phaseoleae*), yet both these strains had *nifD* sequences that grouped with isolates from more distantly related legumes sharing the same geographic region. The

Australian strains from the legume tribe *Bossiaeeae* (*Bossiaea*, *Goodia*) and tribe *Mirbelieae* (*Mirbelia*, *Podolobium*, *Oxylobium*, *Daviesia*) also did not form distinct groups within the *nifD* tree (Fig. 2).

The *nifD* locus occurs sufficiently near to the common *nod* gene cluster (~220 kb apart on the *B. japonicum* USDA 110 chromosome; Kundig *et al.*, 1993) that evolution of host specificity conditioned by *nod* genes may affect *nifD* sequence variation. For example, a 500 kb transmissible chromosomal symbiotic element ('symbiosis island') has been characterized in *Mesorhizobium loti* (Sullivan & Ronson, 1998). If similar transmissible elements existed in *Bradyrhizobium*, then *nod* gene transfer events associated with selection for symbiotic compatibility with novel legume hosts (Sullivan *et al.*, 1995) may also redistribute *nifD* sequences across bacterial lineages. It will be important in future work to clarify mechanisms of gene transfer that have acted in *Bradyrhizobium*.

The prevalence of lateral transfer is an important issue both for bacterial population genetics and systematics. In the presence of lateral transfer, the true genealogical history will vary among genes, or among different portions of the same gene (Dykhuizen & Green, 1991; Zhou *et al.*, 1997; Spratt & Maiden, 1999). Within the genus *Bradyrhizobium*, the basic division between the two primary taxa *B. japonicum* and *B. elkanii* is widely accepted (Kuykendall *et al.*, 1992; Young & Haukka, 1996; Barrera *et al.*, 1997; van Berkum & Fuhrmann, 2000; Willems *et al.*, 2001). However, most existing studies of *Bradyrhizobium* have focused mainly on ribosomal gene sequences. Our *nifD* results provide no support for the interpretation that the genus *Bradyrhizobium* is divided into two primary lineages affiliated with *B. japonicum* or *B. elkanii*, and instead suggest that the genus may be composed of an unknown number of regionally endemic lineages. To avoid biases associated with the idiosyncratic genealogical history of any one gene, it will be important in future work to extend analyses to additional loci (of both housekeeping and symbiotic genes) to provide a broad comparative basis for understanding *Bradyrhizobium* evolution.

ACKNOWLEDGEMENTS

We are grateful to Joanne Pfeil and Patrick Elia for assistance with sequencing, to two anonymous reviewers for comments on the manuscript and to the NSF (grant DEB-9707697) for financial support.

REFERENCES

- Barrera, L. L., Trujillo, M. E., Goodfellow, M., Garcia, F. J., Hernandez-Lucas, I., Davila, G., van Berkum, P. & Martinez-Romero, E. (1997). Biodiversity of bradyrhizobia nodulating *Lupinus* spp. *Int J Syst Bacteriol* **47**, 1086–1091.
- Clark, M. A., Moran, N. A. & Baumann, P. (1999). Sequence evolution in bacterial endosymbionts having extreme base composition. *Mol Biol Evol* **16**, 1586–1598.
- Cunningham, C. W. (1997). Can three incongruence tests predict when data should be combined? *Mol Biol Evol* **14**, 733–740.

- Doyle, J. J. & Davis, J. I. (1998). Homology in molecular phylogenetics: a parsimony perspective. In *Molecular Systematics of Plants. II. DNA Sequencing*, pp. 101–131. Edited by D. E. Soltis, P. S. Soltis & J. J. Doyle. Boston: Kluwer.
- Doyle, J. J., Doyle, J. L., Ballenger, J. A., Dickson, E. E., Kajita, T. & Ohashi, H. (1997). A phylogeny of the chloroplast gene *rbcL* in the Leguminosae: taxonomic correlations and insights into the evolution of nodulation. *Am J Bot* **84**, 541–554.
- Dykhuizen, D. E. & Green, L. (1991). Recombination in *Escherichia coli* and the definition of biological species. *J Bacteriol* **173**, 7257–7268.
- Eardly, B. D., Wang, F.-S. & van Berkum, P. (1996). Corresponding 16S rRNA gene segments in Rhizobiaceae and *Aeromonas* yield discordant phylogenies. *Plant Soil* **186**, 69–74.
- Farris, J. S., Kallersjo, M., Kluge, A. G. & Bult, C. (1995). Testing significance of incongruence. *Cladistics* **10**, 315–319.
- Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783–791.
- Gottfert, M. S., Rothlisberger, S., Kundig, 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.
- Hasegawa, M., Kishino, H. & Yano, T. (1985). Dating the human-ape split by a molecular clock of mitochondrial DNA. *J Mol Evol* **22**, 160–174.
- Haukka, K., Lindstrom, K. & Young, J. P. W. (1998). 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**, 419–426.
- Klappenbach, J. A., Dunbar, J. M. & Schmidt, T. M. (2000). rRNA operon copy number reflects ecological strategies of bacteria. *Appl Environ Microbiol* **66**, 1328–1333.
- Kundig, C., Hennecke, H. & Gottfert, M. (1993). Correlated physical and genetic map of the *Bradyrhizobium japonicum* 110 genome. *J Bacteriol* **175**, 613–622.
- Kundig, C., Beck, C., Hennecke, H. & Gottfert, M. (1995). A single rRNA gene region in *Bradyrhizobium japonicum*. *J Bacteriol* **177**, 5151–5154.
- Kuykendall, D., Saxena, B., Devine, T. E. & Udell, S. E. (1992). Genetic diversity in *Bradyrhizobium japonicum* Jordan 1982 and a proposal for *Bradyrhizobium elkanii* sp. nov. *Can J Microbiol* **38**, 501–505.
- Lafay, B. & Burdon, J. J. (1998). Molecular diversity of rhizobia occurring on native shrubby legumes in southeastern Australia. *Appl Environ Microbiol* **64**, 3989–3997.
- Lafay, B. & Burdon, J. J. (2001). Small-subunit rRNA genotyping of rhizobia nodulating Australian *Acacia* spp. *Appl Environ Microbiol* **67**, 396–402.
- 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.
- Maddison, W. P. & Slatkin, M. (1991). Null models for the number of evolutionary steps in a character on a phylogenetic tree. *Evolution* **45**, 1184–1197.
- Parker, M. A. (1999). Relationships of bradyrhizobia from the legumes *Apios americana* and *Desmodium glutinosum*. *Appl Environ Microbiol* **65**, 4914–4920.
- Parker, M. A. (2000). Divergent *Bradyrhizobium* symbionts on *Tachigali versicolor* from Barro Colorado Island, Panama. *Syst Appl Microbiol* **23**, 585–590.
- Parker, M. A. (2001). Case of localized recombination in 23S rRNA genes from divergent *Bradyrhizobium* lineages associated with neotropical legumes. *Appl Environ Microbiol* **67**, 2076–2082.
- Parker, M. A. & Lunk, A. (2000). Relationships of bradyrhizobia from *Platypodium* and *Machaerium* (Papilionoideae tribe Dalbergieae) on Barro Colorado Island, Panama. *Int J Syst Evol Microbiol* **50**, 1179–1186.
- Qian, J. & Parker, M. A. (2002). Contrasting *nifD* and ribosomal gene relationships among *Mesorhizobium* from *Lotus oroboides* in northern Mexico. *Syst Appl Microbiol* (in press).
- Rigottier-Gois, L., Turner, S. L., Young, J. P. W. & Amarger, N. (1998). Distribution of repC plasmid-replication sequences among plasmids and isolates of *Rhizobium leguminosarum* bv. *viciae* from field populations. *Microbiology* **144**, 771–780.
- Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**, 406–425.
- Shimodaira, H. & Hasegawa, M. (1999). Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol Biol Evol* **16**, 1114–1116.
- Slatkin, M. & Maddison, W. P. (1989). A cladistic measure of gene flow inferred from the phylogenies of alleles. *Genetics* **123**, 603–613.
- Spratt, B. G. & Maiden, M. C. J. (1999). Bacterial population genetics, evolution and epidemiology. *Philos Trans R Soc Lond B* **354**, 701–710.
- Sullivan, J. T. & Ronson, C. W. (1998). Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into a phe-tRNA gene. *Proc Natl Acad Sci USA* **95**, 5145–5149.
- Sullivan, J. T., Patrick, H. N., Lowther, W. L., Scott, D. B. & Ronson, C. W. (1995). Nodulating strains of *Rhizobium loti* arise through chromosomal symbiotic gene transfer in the environment. *Proc Natl Acad Sci USA* **92**, 8885–8889.
- Sullivan, J. T., Eardly, B. D., van Berkum, P. & Ronson, C. W. (1996). Four unnamed species of nonsymbiotic rhizobia isolated from the rhizosphere of *Lotus corniculatus*. *Appl Environ Microbiol* **62**, 2818–2825.
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673–4680.
- Turner, S. L. & Young, J. P. W. (2000). The glutamine synthetases of rhizobia: phylogenetics and evolutionary implications. *Mol Biol Evol* **17**, 309–319.
- Ueda, T., Suga, Y., Yahiro, N. & Matsuguchi, T. (1995). Phylogeny of Sym plasmids of rhizobia by PCR-based sequencing of a *nodC* segment. *J Bacteriol* **177**, 468–472.
- van Berkum, P. & Fuhrmann, J. J. (2000). Evolutionary relationships among the soybean bradyrhizobia reconstructed from 16S rRNA gene and internally transcribed spacer region sequence divergence. *Int J Syst Evol Microbiol* **50**, 2165–2172.
- van Berkum, P., Ruihua, F., Campbell, T. A. & Eardly, B. D. (1998). Some issues of relevance in the taxonomy of rhizobia. In *Proceedings of the 16th North American Conference on Symbiotic Nitrogen Fixation*, pp. 1–5. Cancun, Mexico.
- Wang, Y. & Zhang, Z. (2000). 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**, 2845–2854.
- Wendel, J. F. & Doyle, J. J. (1998). Phylogenetic incongruence: window into genome history and molecular evolution. In *Molecular Systematics of Plants. II. DNA Sequencing*, pp.

- 265–296. Edited by D. E. Soltis, P. S. Soltis & J. J. Doyle. Boston: Kluwer.
- Wernegreen, J. J. & Riley, M. A. (1999).** Comparison of the evolutionary dynamics of symbiotic and housekeeping loci: a case for the genetic coherence of rhizobial lineages. *Mol Biol Evol* **16**, 98–113.
- Wernegreen, J. J., Harding, E. E. & Riley, M. A. (1997).** Rhizobium gone native: unexpected plasmid stability of indigenous *Rhizobium leguminosarum*. *Proc Natl Acad Sci U S A* **94**, 5483–5488.
- Werren, J. H., Zhang, W. & Guo, L. R. (1995).** Evolution and phylogeny of *Wolbachia*: reproductive parasites of arthropods. *Proc R Soc Lond B* **261**, 55–71.
- Willems, A., Coopman, R. & Gillis, M. (2001).** Phylogenetic and DNA-DNA hybridization analyses of *Bradyrhizobium* species. *Int J Syst Evol Microbiol* **51**, 111–117.
- Wilson, J. K. (1934).** Longevity of *Rhizobium japonicum* in relation to its symbiont on the soil. *Cornell Univ Agric Exp Stn Mem* **162**, 1–11.
- Woese, C. R. (1987).** Bacterial evolution. *Microbiol Rev* **51**, 221–271.
- Xu, L. M., Ge, C., Cui, Z., Li, J. & Fan, H. (1995).** *Bradyrhizobium liaoningense* sp. nov., isolated from the root nodules of soybeans. *Int J Syst Bacteriol* **45**, 706–711.
- Yap, W. H., Zhang, Z. & Wang, Y. (1999).** Distinct types of rRNA operons exist in the genome of the actinomycete *Thermomonospora chromogena* and evidence for horizontal transfer of an entire rRNA operon. *J Bacteriol* **181**, 5201–5209.
- Young, J. P. W. & Wexler, M. (1988).** Sym plasmid and chromosomal genotypes are correlated in field populations of *Rhizobium leguminosarum*. *J Gen Microbiol* **134**, 2731–2739.
- Young, J. P. W. & Haukka, K. E. (1996).** Diversity and phylogeny of rhizobia. *New Phytol* **133**, 87–94.
- Zhou, J., Bowler, L. D. & Spratt, B. G. (1997).** Interspecies recombination, and phylogenetic distortions, within the glutamine synthetase and shikimate dehydrogenase genes of *Neisseria meningitidis* and commensal *Neisseria* species. *Mol Microbiol* **23**, 799–812.
-
- Received 12 March 2002; revised 17 April 2002; accepted 7 May 2002.