Article:
Gaunt, M.W., Rigottier-Gois, L., Lloyd-Macgilp, S.A. et al. (2 more authors) (2001) Phylogenies of atpD and recA support the small subunit rRNA-based classification of rhizobia. International Journal of Sytematic and Evolutionary Microbiology. pp. 2037-2048. ISSN 1466-5034

International Journal of Systematic and Evolutionary Microbiology (2001), **51**, 2037–2048 Printed in Great Britain

Phylogenies of *atpD* and *recA* support the small subunit rRNA-based classification of rhizobia

Department of Biology, University of York, PO Box 373, York YO10 5YW, UK

M. W. Gaunt,† S. L. Turner,‡ L. Rigottier-Gois,§ S. A. Lloyd-Macgilp‖ and J. P. W. Young

Author for correspondence: J. P. W. Young. Tel: +44 1904 432914. Fax: +44 1904 432860. e-mail: jpy1@york.ac.uk

The current classification of the rhizobia (root-nodule symbionts) assigns them to six genera. It is strongly influenced by the small subunit (16S, SSU) rRNA molecular phylogeny, but such single-gene phylogenies may not reflect the evolution of the genome as a whole. To test this, parts of the *atpD* and *recA* genes have been sequenced for 25 type strains within the *α-Proteobacteria*, representing species in *Rhizobium*, *Sinorhizobium*, *Mesorhizobium*, *Bradyrhizobium*, *Azorhizobium*, *Agrobacterium*, *Phyllobacterium*, *Mycoplana* and *Brevundimonas*. The current genera *Sinorhizobium* and *Mesorhizobium* are well supported by these genes, each forming a distinct phylogenetic clade with unequivocal bootstrap support. There is good support for a *Rhizobium* clade that includes *Agrobacterium tumefaciens*, and the very close relationship between *Agrobacterium rhizogenes* and *Rhizobium tropici* is confirmed. There is evidence for recombination within the genera *Mesorhizobium* and *Sinorhizobium*, but the congruence of the phylogenies at higher levels indicates that the genera are genetically isolated. rRNA provides a reliable distinction between genera, but genetic relationships within a genus may be disturbed by recombination.

Keywords: *Rhizobium*, *Sinorhizobium*, *Mesorhizobium*, *recA*, *atpD*

INTRODUCTION

The rhizobia are root-nodulating bacteria responsible for a significant part of the global fixation of nitrogen. The ability of rhizobia to nodulate plants and reduce N_2 is conferred by genes that are plasmid-borne in many species (Pueppke, 1996), and the lateral transfer of these genes is the most likely explanation for their occurrence within several distinct clades of subgroup 2 of the *α-Proteobacteria* (Dobert *et al.*, 1994; Kaijalainen & Lindström, 1989; Lindström *et al.*, 1995; Young, 1998; Young & Johnston, 1989). The phylogeny of the nodulation genes is quite different from that of the small subunit rRNA (SSU or 16S rRNA) genes in these bacteria. As in other bacterial groups, the SSU phylogeny has had a major influence on our current perception of evolutionary relationships among rhizobia (Willems & Collins, 1993; Young, 1996; Young *et al.*, 1991). More than 20 species have been described, and they are classified into the genera *Rhizobium*, *Sinorhizobium*, *Allorhizobium*, *Mesorhizobium*, *Bradyrhizobium* and *Azorhizobium* as well as '*Methylobacterium nodulans*' (de Lajudie *et al.*, 1994, 1998b; Dreyfus *et al.*, 1988; Jarvis *et al.*, 1997; Jordan, 1982; Sy *et al.*, 2001). For current nomenclature, see J. P. Euzéby's *List of Bacterial Names with Standing in Nomenclature* (http://www.bacterio.cict.fr/).

The genera *Bradyrhizobium* and *Azorhizobium* have many phenotypic differences from the others, includ-

†**Present address:** Parasite Molecular Biology and Biochemistry Unit, Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK.

‡**Present address:** CEH Oxford, Mansfield Road, Oxford OX1 3SR, UK.

§**Present address:** Unité d'Ecologie et Physiologie du Système Digestif, INRA Centre de Recherche de Jouy-en-Josas, 78352 Jouy-en-Josas Cedex, France.

‖**Present address:** Molecular Medicine Centre, University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, UK.

The *atpD* and *recA* sequences and details of the SimPlot analyses are available as supplementary material in IJSEM Online (http://ijs.sgmjournals.org/).

Abbreviations: HKY85, Hasegawa–Kishino–Yano model; K2P, Kimura's two-parameter model; SSU, small subunit rRNA.

The EMBL accession numbers for the sequences reported in this study are AJ294386–AJ294409 (*atpD*) and AJ294363–AJ294385 (*recA*).

ing much slower growth on laboratory media (Dreyfus *et al*., 1988; Fred *et al*., 1932; Jordan, 1982). They are also fairly distant in the SSU phylogeny, being no more related to the other rhizobia, or to each other, than are many non-symbiotic bacteria such as *Beijerinckia indica* or *Blastochloris* (*Rhodopseudomonas*) *viridis* (Young, 1996; Young & Haukka, 1996). '*Methylobacterium nodulans*' is the name suggested for a recently discovered fourth group of root-nodule bacteria that are methylotrophic and closely related to known *Methylobacterium* species (Sy *et al*., 2001). The other rhizobia, known collectively as the fast-growers, are more closely related, but their clade in the SSU phylogeny (Maidak *et al*., 2000) also includes *Agrobacterium* (plant gall-formers), *Phyllobacterium* (leaf and rhizosphere colonists), *Mycoplana* (branching soil bacteria), *Brevundimonas bullata* (formerly *Mycoplana bullata*, see Abraham *et al*., 1999) and possibly the animal pathogens *Bartonella* and *Brucella* (Moreno *et al*., 1990; Weisburg *et al*., 1985; Young *et al*., 1991).

There has been controversy over the relationship between *Agrobacterium* and *Rhizobium*, which have long been recognized as close relatives (Fred *et al*., 1932). Some authors suggested that *Agrobacterium* should be subsumed into *Rhizobium* (de Ley, 1968; Graham, 1964, 1976; Heberlein *et al*., 1967; Kerr, 1992; Sawada *et al*., 1993; White, 1972), but the splitting of *Sinorhizobium* and *Mesorhizobium* from *Rhizobium* weakened the case for such an amalgamation because it was not clear whether *Agrobacterium* formed a monophyletic clade with *Rhizobium sensu stricto*. In the SSU phylogeny, *Rhizobium leguminosarum*, *Rhizobium etli* and *Rhizobium tropici* form a clade, recently joined by *Rhizobium gallicum* (Amarger *et al*., 1997) and *Rhizobium mongolense* (van Berkum *et al*., 1998). This clade includes *Agrobacterium rhizogenes*, whose SSU is so similar to that of *Rhizobium tropici* that one might suggest they were conspecific (Young & Haukka, 1996). A second clade is formed by *Agrobacterium tumefaciens* and *Agrobacterium rubi* and, in some SSU phylogenies, this is joined peripherally by *Agrobacterium vitis*, *Rhizobium galegae* and, more recently, *Rhizobium huautlense* (Wang *et al*., 1998) and *Allorhizobium undicola* (de Lajudie *et al*., 1998b). Young *et al*. (2001) have recently proposed the inclusion of *Agrobacterium* and *Allorhizobium* in *Rhizobium*. This proposal is supported by a maximum-likelihood analysis of the SSU data that provides stronger evidence for the monophyly of this expanded genus *Rhizobium* and it sidesteps the necessity to resolve the phylogenetic uncertainties within it. It does, however, create a genus that has much greater SSU diversity than any of the other rhizobial genera. For clarity, we have retained the separate names in this paper.

The enormous success of SSU sequencing as an approach to bacterial systematics relies on the assumption that the evolutionary relationships of bacteria are reliably reflected in their rRNA genes, but gene trees are not necessarily an accurate representation of the species tree if there is exchange of chromosomal genes among the bacteria in question (Pamilo & Nei, 1988). It is clearly desirable to examine genes other than SSU in order to confirm the relationships among the rhizobia and to clarify those that are ambiguous in the SSU phylogeny. We recently presented phylogenies for some species of fast-growing rhizobia (but not agrobacteria or other non-rhizobia) based on the two glutamine synthetase (GS) loci (Turner & Young, 2000). The phylogeny of the GSI gene is very similar to that of SSU, with well-supported clades for *Mesorhizobium*, *Sinorhizobium* and *Rhizobium* (*Rhizobium leguminosarum*, *Rhizobium etli* and *Rhizobium tropici*), while *Rhizobium galegae* occupies a position intermediate between *Rhizobium* and *Sinorhizobium*. The GSII phylogeny is similar in most respects, but *Rhizobium galegae* is closely allied with *Rhizobium leguminosarum* and *Rhizobium etli*, *Mesorhizobium huakuii* is closer to *Rhizobium* than to *Mesorhizobium* and the *Bradyrhizobium* sequence is much closer to those of the fast-growing rhizobia than one would expect. We think these discrepancies probably reflect gene transfer.

In order to assess the impact of gene transfer on the interpretation of single-gene phylogenies, such as those based on SSU, it is necessary to study more loci. We have now examined two additional loci, *atpD* and *recA*, in the type strains of representative species of fast-growing rhizobia and their relatives. The *recA* gene encodes part of the DNA recombination and repair system and large-scale phylogenies of *recA* in the bacteria are consistent with the corresponding SSU phylogenies (Eisen, 1995; Young, 1998). The *atpD* gene encodes the beta subunit of the membrane ATP synthase that is essential for energy production. It is a ubiquitous and conserved gene, but there are no published sequences from rhizobia. In the complete 7036071 bp chromosome sequence of *Mesorhizobium loti* strain MAFF303099 (Kaneko *et al*., 2000), *atpD* starts at position 3246882, *recA* at 24004 and the two SSU genes at 2745482 and 2752970. The genes are therefore spaced well apart on the genome.

METHODS

Isolates. Type strains of rhizobia, *Agrobacterium*, *Phyllobacterium*, *Brevundimonas* and *Mycoplana* species were obtained as detailed in Table 1. Cultures were grown at 28 °C in TY medium (Beringer, 1974) and the medium was solidified as required with 1% (w/v) agar.

Molecular methods. DNA was purified from liquid cultures as described by Laguerre *et al*. (1992). Oligonucleotide primers for *recA* and *atpD* were designed from alignments of published nucleotide sequences and more robust primers were subsequently designed on the basis of the early sequences obtained. The primers are described in Table 2 and were used for both PCR and sequencing reactions. Purified DNA was added to a 50 µl PCR mixture [200 mM dNTP, 1·5 mM $MgCl_2$, 10 pmol each primer, 1 U *Taq* DNA polymerase and 1× reaction buffer (Promega)] and incubated at 95 °C for 5 min followed by 30 rounds of thermal cycling (94 °C for 45 s, 50 °C for 60 s and 74 °C for 90 s in a

Table 1. Bacterial strains and sequences used in this study

Source of strains: USDA, USDA–ARS National Rhizobium Germplasm Collection, Soybean and Alfalfa Research Laboratory, Beltsville, MD, USA; NCPPB, National Collection of Plant Pathogenic Bacteria, Central Science Laboratory, MAFF, Sand Hutton, York, UK; HAMBI, The Culture Collection, Department of Applied Chemistry and Microbiology, University of Helsinki, Finland; A321^T, Sophie Rome, Laboratoire d'Ecologie Microbienne du Sol UMR CNRS 5557, Villeurbanne, France.

Species	Type strain	Sequence accession number*		
		SSU	*atpD*	*recA*
Rhizobium etli	USDA 9032^T	U28916[a]	**AJ294404**	**AJ294375**
Rhizobium galegae	USDA 4128^T	X67226[b]	**AJ294406**	**AJ294378**
Rhizobium leguminosarum bv. *viciae*	USDA 2370^T	U29386[a]	**AJ294405**	**AJ294376**
Rhizobium tropici A	USDA 9039^T	X67233[b]	**AJ294396**	**AJ294372**
Rhizobium tropici B	USDA 9030	U89832[c]	**AJ294397**	**AJ294373**
Agrobacterium rhizogenes	NCPPB 2991^T	X67224[b]	**AJ294398**	**AJ294374**
Agrobacterium tumefaciens	NCPPB 2437^T	X67223[b]	**AJ294407**	**AJ294377**
Sinorhizobium fredii	USDA 205^T	X67231[b]	**AJ294402**	**AJ294379**
Sinorhizobium meliloti	USDA 1002^T	X67222[b]	**AJ294400**	**AJ294382**
Sinorhizobium medicae	A321^T	L39882[d]	**AJ294401**	**AJ294381**
Sinorhizobium sahelense	HAMBI 217^T	X68390[e]	**AJ294399**	**AJ294380**
Sinorhizobium terangae	HAMBI 220^T	X68388[e]	**AJ294403**	**AJ294383**
Mesorhizobium ciceri	USDA 3383^T	U07934[f]	**AJ294395**	**AJ294367**
Mesorhizobium huakuii	USDA 4779^T	D12797[g]	**AJ294394**	**AJ294370**
Mesorhizobium loti	USDA 3471^T	X67229[b]	**AJ294393**	**AJ294371**
Mesorhizobium mediterraneum	USDA 3392^T	L38825[h]	**AJ294391**	**AJ294369**
Mesorhizobium tianshanense	USDA 3592^T	AF041447[i]	**AJ294392**	**AJ294368**
Bradyrhizobium japonicum	USDA 6^T	U69638[j]	**AJ294388**	–
Azorhizobium caulinodans	USDA 4892^T	X67221[b]	**AJ294389**	**AJ294363**
Phyllobacterium rubiacearum	ATCC 43591^T	D12790[g]	**AJ294386**	**AJ294366**
Phyllobacterium myrsinacearum	ATCC 43590^T	D12789[g]	**AJ294387**	**AJ294365**
Mycoplana dimorpha	ATCC 4279^T	D12786[g]	**AJ294408**	**AJ294385**
Mycoplana ramosa	ATCC 49678^T	D10463[k]	**AJ294409**	**AJ294384**
Brevundimonas bullata	ATCC 4278^T	D12785[g]	**AJ294390**	**AJ294364**
Rhodobacter capsulatus		D16428[l]	X99599[q]	X82183[v]
Rhodobacter sphaeroides		X53853[m]	–	X72705[w]
Rhodobacter blasticus		–	Z00018[r]	–
Rickettsia prowazekii		M21789[n]	AJ235273[s]	U01959[x]
Helicobacter pylori		U01330[o]	AF004014[t]	U13756[y]
Pirellula marina		X62912[p]	X57204[u]	–

* Sequences in bold were determined in this study. Sources of published sequences: *a*, van Berkum *et al*. (1996); *b*, Willems & Collins (1993); *c*, van Berkum *et al*. (1998); *d*, Rome *et al*. (1996); *e*, de Lajudie *et al*. (1994); *f*, Nour *et al*. (1994); *g*, Yanagi & Yamasato (1993); *h*, Nour *et al*. (1995); *i*, Wang *et al*. (1999); *j*, Barrera *et al*. (1997); *k*, Sakurada *et al*. (1996); *l*, Hiraishi & Ueda (1994); *m*, Dryden & Kaplan (1990); *n*, Weisburg *et al*. (1989); *o*, Eckloff *et al*. (1994); *p*, Liesack *et al*. (1992); *q*, Borghese *et al*. (1998); *r*, Tybulewicz *et al*. (1984); *s*, Andersson *et al*. (1998); *t*, McGowan *et al*. (1997); *u*, Ronner *et al*. (1991); *v*, Fernandez de Henestrosa *et al*. (1995); *w*, Calero *et al*. (1994); *x*, Dunkin & Wood (1994); *y*, Thompson & Blaser (1995).

PTC-100 thermocycler; M J Research, Inc.). PCR products were evaluated on a 1% agarose gel stained with ethidium bromide. Unincorporated primers and dNTPs were removed from PCR products with Qiaquick PCR purification kits (Qiagen) or, if necessary, products were gel-purified and extracted with Qiagen kits. Purified DNA was sequenced using the dideoxynucleotide chain-termination method with fluorescent ddNTPs from Applied Biosystems on an ABI 377 Prism automated DNA sequencer, in accordance with the manufacturer's instructions. Faint PCR products were cloned into a sequencing vector (pGEM-T; Promega) and transformed into competent *Escherichia coli* XL-1 Blue before sequencing using T7 and SP6 primers. To check the location of the genes, plasmids were separated by the Eckhardt method of in-well lysis (Eckhardt, 1978; Wheatcroft *et al*., 1990), transferred to nylon membranes and hybridized with PCR products of each gene amplified from *Rhizobium leguminosarum* USDA 2370^T using the DIG detection kit (Boehringer).

Phylogenetic analysis. Nucleotide alignments of partial *recA*, *atpD* and SSU sequences were assembled using CLUSTAL X (Thompson *et al*., 1997) and optimized by hand. Molecular phylogenies were reconstructed using the neighbour-joining method (Saitou & Nei, 1987), with distances estimated using Jukes–Cantor, Kimura's two-parameter (K2P), Galtier and Gouy, Hasegawa–Kishino–Yano (HKY85), Log-det and

Table 2. Primers used for PCR and sequencing

Gene	Position*	Direction	Primer length (bases)	Primer sequence (5′–3′)†
recA	6	Forward	26	CGK CTS GTA GAG GAY AAA TCG GTG GA
recA	63	Forward	23	ATC GAG CGG TCG TTC GGC AAG GG
recA	504	Reverse	20	TTG CGC AGC GCC TGG CTC AT
recA	555	Reverse	26	CGR ATC TGG TTG ATG AAG ATC ACC AT
atpD	273	Forward	21	SCT GGG SCG YAT CMT GAA CGT
atpD	294	Forward	20	ATC GGC GAG CCG GTC GAC GA
atpD	771	Reverse	23	GCC GAC ACT TCC GAA CCN GCC TG

* First base amplified after the primer, relative to the start of the *recA* gene of *Rhizobium leguminosarum* (X59956) or the *atpD* gene of *Rhodobacter capsulatus* (X99599).

† Mixtures of bases used at certain positions are given as: K, T or G; S, G or C; Y, C or T; R, A or G; M, A or C; N, any base.

Tajima and Nei models using CLUSTAL X, PHYLOWIN (Galtier *et al.*, 1996) and PAUP*4.0 (Swofford, 1997), with 1000 bootstrap replications.

Maximum-likelihood analysis was performed using three different software implementations. In fastDNAml (Olsen *et al.*, 1994), which uses the F84 model of nucleotide change, global searches were combined with a reiterative jumble of the taxon order and each search was repeated until three identical likelihoods were obtained. The genetic algorithm program GAML (Lewis, 1998) uses the HKY85 model and was run with 10 individuals, with two copies of the fittest always held over, and crossover, branch-length mutation and topological mutation probabilities all set to 0·4. Finally, PAUP*4.0 (Swofford, 1997) was used with the HKY85 model incorporating a discrete gamma (Γ) distribution of site-specific rates and heuristic searches were conducted using the tree bisection–reconnection method for more than 1500 branch rearrangements. Maximum-likelihood estimates of transition/transversion ratios were obtained by a manual optimization using HKY85. Base frequencies and the proportion of invariant sites were estimated directly from empirical data. The estimate of the shape parameter, α, of the Γ distribution for three discrete rate categories was obtained through an automated maximum-likelihood search strategy incorporating all *a priori* parameters described.

The χ^2 test of homogeneity of base frequencies, *p*-distances, the skewness statistic of tree length distributions (10000 random trees) and the partition homogeneity test (1000 random partitions) were calculated using PAUP*4.0. Kishino–Hasagawa tests were performed using fastDNAml and PAUP*4.0. Alternative phylogeny branching orders were obtained using MacClade (Maddison & Maddison, 1992).

The Windows program SimPlot (S. C. Ray; http://www.med.jhu.edu/deptmed/sray/download/) was used to plot pairwise sequence similarities for a sliding window along the alignments and for bootscanning, that is, plotting bootstrap support for pairings of taxa in neighbour-joining trees using a sliding window (Salminen *et al.*, 1995).

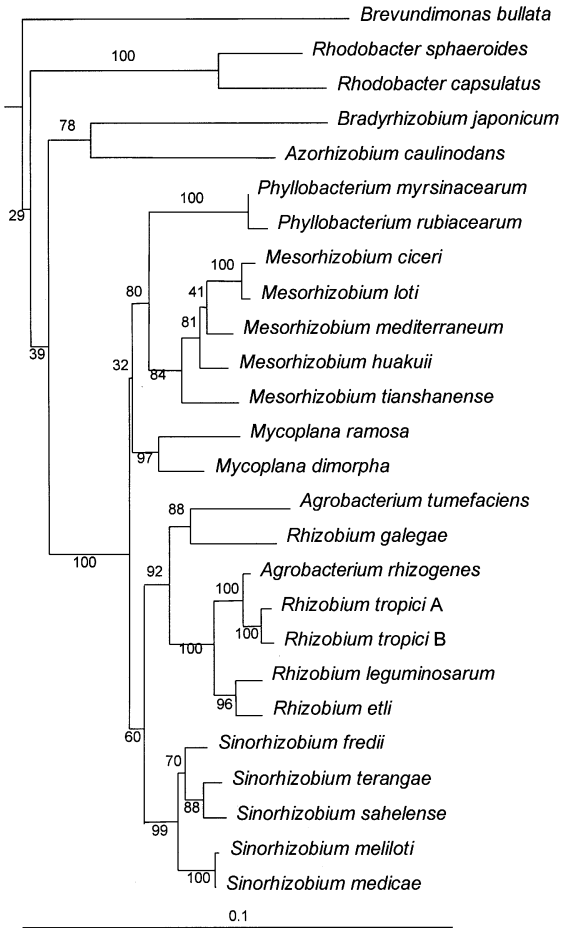

Fig. 1. Neighbour-joining phylogeny for SSU (16S rRNA) genes of type strains of rhizobia and relatives, based on published sequences (Table 1). The tree is rooted using the sequence from *Rickettsia prowazekii* (not shown). Alignment length was 1400 bp, positions with gaps were omitted, K2P distance correction applied. Percentage bootstrap support (1000 replicates) is shown for each internal branch.

RESULTS

DNA sequencing and gene location

PCR amplification and sequencing of rhizobium type strains provided at least 450 bp of reliable *atpD* sequence for all 25 type strains in the analysis and 441 bp of reliable *recA* sequence for 24 type strains. These sequences are available as supplementary data in IJSEM Online (http://ijs.sgmjournals.org/). Repeated attempts to amplify and sequence *Bradyrhizobium japonicum recA* were unsuccessful. The

reasons for this are not known. The alignments used for phylogenetic analysis corresponded to nucleotides 301–746 in the *Rhodobacter capsulatus atpD* coding sequence (accession no. X99599) and 75–515 in the *Rhizobium leguminosarum recA* coding sequence (X59956). The alignment lengths were 461 bp for *atpD* and 441 bp for *recA*. Labelled probes prepared from the *atpD* and *recA* genes of *Rhizobium leguminosarum* hybridized to sheared DNA of each of the strains on Eckhardt plasmid gels but not to plasmid bands (data not shown). We conclude that all the genes we sequenced are chromosomal.

Data robustness

All *atpD*, *recA* and SSU datasets and subsets were assessed using the χ^2 homogeneity test of base frequencies. Only the extensively rooted ('full') *recA* dataset described below showed significant variation ($P < 0{\cdot}01$, with $P > 0{\cdot}39$ in all other cases). All phylogenies also showed *p*-distances below 75% before and after the removal of invariant sites. Furthermore, phylogenies constructed using only third codon positions were broadly congruent with those produced using all three codon positions. Finally, all datasets showed highly significant phylogenetic structure ($P < 0{\cdot}01$; $g_1 = -0{\cdot}63$ for 148 informative sites, $-0{\cdot}6216$ for 214 informative sites and $-0{\cdot}667$ for 392 informative sites for full datasets of *recA*, *atpD* and SSU, respectively) using the skewness statistic of tree length distribution (Hillis & Huelsenbeck, 1992).

The datasets therefore show strong phylogenetic signal and provide no evidence of saturation. Although variation in base frequency was observed in *recA*, the distant outgroups were maintained for the preliminary calculations to ensure consistent rooting.

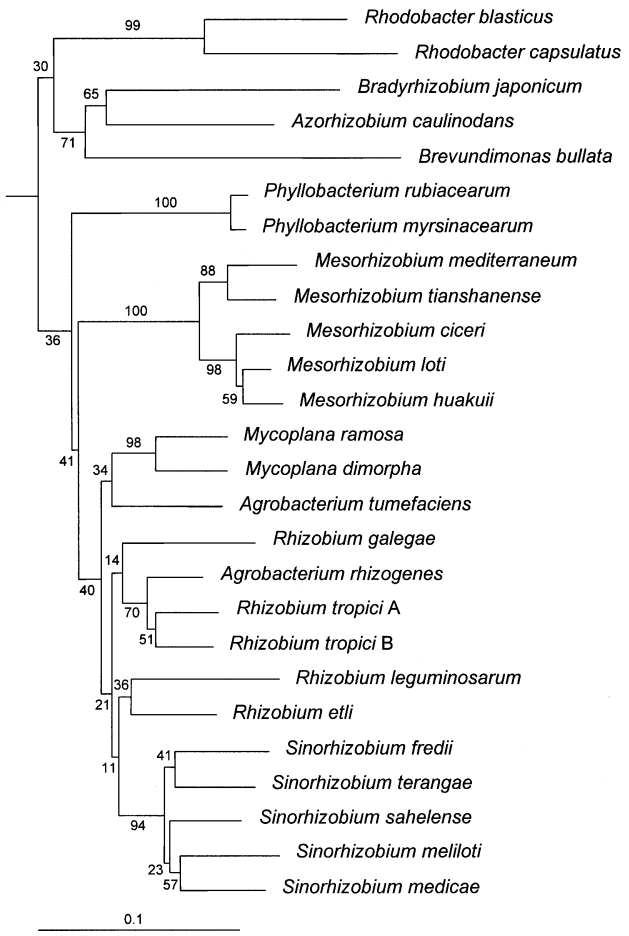

Fig. 2. Neighbour-joining phylogeny for partial *atpD* sequences (alignment length 461 bp) of type strains of rhizobia and relatives. Other details as Fig. 1.

Neighbour-joining phylogeny

Neighbour-joining phylogenies of *atpD* and *recA* were constructed for the 25 type strains sequenced in this study (*Rhizobium*, *Sinorhizobium*, *Mesorhizobium*, *Bradyrhizobium*, *Azorhizobium*, *Agrobacterium*, *Phyllobacterium*, *Brevundimonas* and *Mycoplana* species) plus published outgroup sequences and for SSU of the same species using published sequences. Each phylogeny was rooted using *Helicobacter pylori* and (for SSU and *atpD*) *Pirellula marina*, both of which are unambiguously external to the α-*Proteobacteria*. For SSU, a more comprehensive phylogeny containing an additional 174 taxa of the α-*Proteobacteria* sharing an ancestral node with *Rhodospirillum* was also constructed. In all cases, *Rhodobacter* sequences were external to the clade that included all rhizobia and *Rickettsia prowazekii* sequences were more distant still. *Rickettsia* was subsequently used as the outgroup and the more remote outgroups were removed. Neighbour-joining phylogenies were constructed for each dataset using the six distance-correction models described in Methods. Those clades that had strong bootstrap support were found consistently using all models and the trees obtained with the K2P model are shown (Figs 1–3), since more complex models gave substantially similar results.

For each of the genes (SSU, *recA*, *atpD*), the five *Mesorhizobium* species formed a strongly supported monophyletic group (84–100% of bootstraps), as did the five species of *Sinorhizobium* (82–99%). Support for *Mesorhizobium* in the SSU tree increased from 84 to 100% if the *Phyllobacterium* sequences were omitted. The status of the other fast-growing rhizobia (the genus *Rhizobium*) was less clear-cut and has to be considered jointly with that of *Agrobacterium*. The close similarity between *Agrobacterium rhizogenes* and *Rhizobium tropici*, previously noted on the basis of SSU and supported here by 100% of bootstraps, was confirmed by both the other genes (99% for *recA*, 70% for *atpD*). In the SSU tree, *Rhizobium leguminosarum* and *Rhizobium etli* were closely related (96%) and formed the strongly supported sister group of the *Rhizobium tropici–Agrobacterium rhizogenes* clade (100%). *Rhizobium galegae* and *Agrobacterium tumefaciens* were associated with these species (92%)

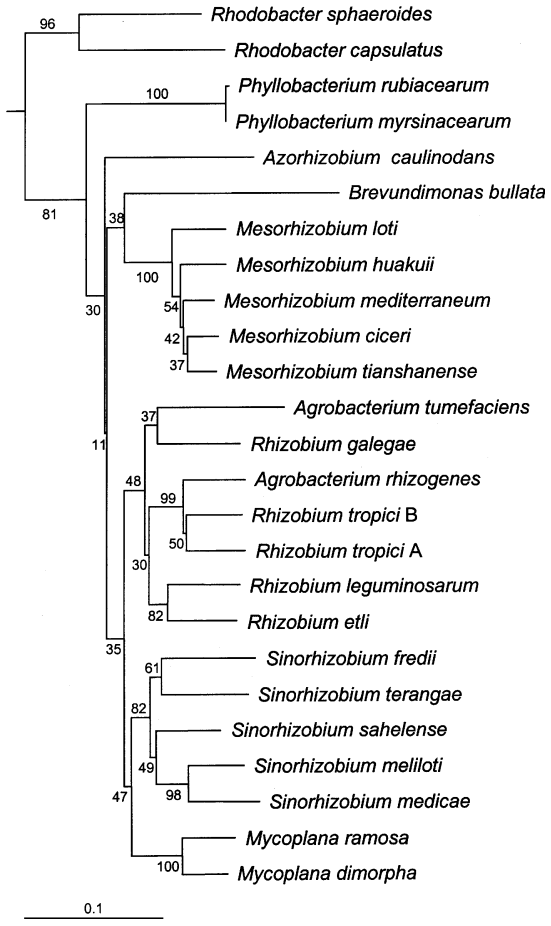

Fig. 3. Neighbour-joining phylogeny for partial *recA* sequences (alignment length 441 bp) of type strains of rhizobia and relatives. Other details as Fig. 1.

and formed a clade together (88%). All these relationships were also seen in the *recA* tree, though with poorer bootstrap support. In the *atpD* tree, bootstrap support was generally poor, and the *Rhizobium–Agrobacterium* clade was not supported, though, on the other hand, it was not excluded either.

The phylogenetic positions of the non-rhizobial genera *Mycoplana*, *Brevundimonas* and especially *Phyllobacterium* were less robust and this contributes markedly to the ambiguity in parts of the trees. *Mycoplana ramosa* and *Mycoplana dimorpha* were sister taxa in the trees for all three genes, with good bootstrap support (97–100%), and formed an independent clade not associated closely with any other genus. For all three genes, the *Brevundimonas bullata* branch was unusually long and its position within the phylogeny was poorly supported.

The two *Phyllobacterium* species, *Phyllobacterium rubiacearum* and *Phyllobacterium myrsinacearum*, were almost identical in sequence at all three loci, but their placement in the trees was uncertain. The association with *Mesorhizobium* seen for SSU (Fig. 1) was not supported by the other genes, which suggested a deeper placement (Figs 2 and 3). Removal of the two *Phyllobacterium recA* sequences did not alter the structure of the remainder of the tree at all, but dramatically improved support for some clades. Notably, support for *Rhizobium–Agrobacterium* rose from 48 to 89%.

The *atpD* tree, on the other hand, which generally had poorer resolution than that for *recA*, was not significantly altered by the removal of *Phyllobacterium* and/or *Brevundimonas* sequences. Removal of *Rickettsia*, the outgroup, had a more substantial effect on the *atpD* tree: support for the crown group (*Phyllobacterium* to *Sinorhizobium*) increased from 36 to 74%, while the root of this group shifted strongly onto the *Mesorhizobium* branch, so that the clade that excludes *Mesorhizobium* (but includes *Phyllobacterium*) received 98% bootstrap support. Other relationships were not significantly affected and the improvements in support occurred whether or not *Phyllobacterium* and/or *Brevundimonas* sequences were included in the tree.

The program SimPlot was used to look for possible reasons for the unstable behaviour of some sequences. Pairwise similarities were plotted for a 100 bp sliding window along the length of each gene and bootscanning analyses were made to highlight changes in support for different tree topologies. For the 5′ part of the *recA* sequence, *Brevundimonas* was associated with *Mesorhizobium* and *Phyllobacterium* lay near the base of the tree, while the situation was reversed in the 3′ part. This explains the low bootstrap values (30, 11, 38%) in Fig. 3. In the SSU, the *Rhizobium galegae* sequence generally resembles that of *Agrobacterium tumefaciens* except for a cluster of phylogenetically informative sites at around position 1000 that match *Rhizobium leguminosarum* instead. This reduced the confidence with which *Rhizobium galegae* could be placed in the tree and might perhaps reflect a recombination event. Details of the analyses are available as supplementary information in IJSEM Online (http://ijs.sgmjournals.org/).

Maximum-likelihood phylogeny

Maximum-likelihood models were constructed as described in Methods. In all cases, the HKY85+Γ model (which allows nucleotide positions to have different evolutionary rates assigned from a gamma distribution) showed a significant improvement on the HKY85 model, according to the likelihood ratio test (Yang *et al.*, 1994). As in the neighbour-joining trees, the position of *Phyllobacterium* was ambiguous for each of the three genes. For the HKY85 model, *Phyllobacterium* formed an outgroup to the fast-growing rhizobia, whilst, for the HKY85+Γ model, it occurred within the fast-growing rhizobia.

Identical topologies between maximum-likelihood and neighbour-joining methods were observed within *Sinorhizobium* for all genes and within the *Rhizobium–Agrobacterium* clade for SSU. Elsewhere,

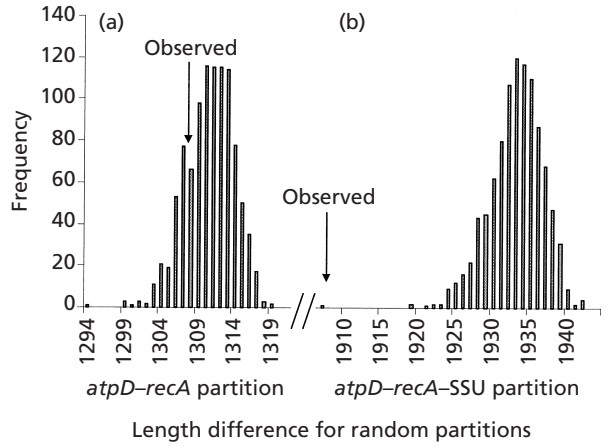

Fig. 4. Partition-homogeneity test to determine whether data from different genes can be combined. The incongruence length difference (the extra steps required when data for genes are combined) is compared with that for 1000 random partitions of the variable sites. (a) Combination of *atpD* and *recA*; (b) combination of *atpD*, *recA* and SSU sequences, for the 23 taxa sequenced in this study.

numerous topological differences between neighbour-joining and maximum-likelihood methods were observed, but none involved branches that were well supported by neighbour-joining bootstrap analysis.

Partition homogeneity and combined data

A more reliable phylogeny might be obtained by using combined data for all three genes. The partition homogeneity test (Cunningham, 1997; Farris *et al.*, 1994; Mickevich & Farris, 1981) was applied to check whether trees for the different genes were sufficiently similar in rates of divergence and branching order that it would be legitimate to combine the data. The test calculates how much longer the combined parsimony tree is than the sum of the separate trees, and compares this difference to the expected distribution estimated by allocating the data randomly into partitions of the same size as the genes. The *atpD* and *recA* sequences for the 23 sequenced type strains (the taxa in Fig. 3 minus *Rickettsia* and *Rhodobacter*) were combined and analysed using a two-partition test, one data partition corresponding to each gene. A three-partition test was conducted on the *atpD*, *recA* and SSU data. Invariant characters were removed for all analyses. Parsimony step-length homogeneity was supported for the *atpD*+*recA* test ($P = 0{\cdot}25$) but rejected for the *atpD*+*recA*+SSU test ($P < 0{\cdot}001$), as shown in Fig. 4. The latter result was confirmed using a two-partition test containing *atpD* and *recA* in one partition and SSU in the second for 100 randomized data partitions ($P < 0{\cdot}01$). These results permit us to combine the *atpD* and *recA* data, but suggest that the SSU data should be treated separately. Similar tests with a reduced dataset of just the 18 rhizobia and agrobacteria led to the same conclusion ($P = 0{\cdot}256$ for *atpD*+*recA*, $P < 0{\cdot}001$ for *atpD*+*recA*+SSU).

Maximum-likelihood and neighbour-joining phylogenies for the combined *atpD* and *recA* sequences were compared with the corresponding SSU phylogenies. Trees were constructed with various combinations of the 23 taxa. Bootstrap support for the *Rhizobium–Agrobacterium* clade was still low (58%) in the neighbour-joining analysis of the combined *atpD*+*recA* data if all taxa were retained, though higher than for either gene separately (48% for *recA*, $<<$ 50% for *atpD*). However, removing non-rhizobia, and especially the troublesome *Phyllobacterium* sequences, increased support to 93%. Trees that include just rhizobia and agrobacteria are shown in Fig. 5.

The robustness of the *Rhizobium–Agrobacterium* clade within the 23-taxon *atpD*+*recA* tree was also investigated by constructing 28 alternative topologies for the disruption of monophyly by moving combinations of *Rhizobium leguminosarum*, *Rhizobium etli* and *Rhizobium galegae*, assessed using the Kishino–Hasagawa test on the HKY85+Γ maximum-likelihood model. This test determines whether an alternative tree is significantly less likely than the best tree, taking into account the observed variance in log-likelihood difference across nucleotide sites (log-likelihood, ln L, is the natural logarithm of the likelihood, so a reduction of 1 in ln L represents a tree that is 2·718 times less likely). The rearrangement of highest likelihood was when the *Rhizobium leguminosarum–Rhizobium etli* lineage formed a sister group to *Sinorhizobium–Mycoplana* ($P = 0{\cdot}15$; reduction in ln L of 8·21). The same placement of *Rhizobium galegae* or the three-taxon lineage also showed no significant reduction in likelihood ($P \leqslant 0{\cdot}0568$). However, placement of *Rhizobium leguminosarum–Rhizobium etli* or *Rhizobium galegae* within the *Sinorhizobium* clade was significantly worse ($P \leqslant 0{\cdot}0369$), as was placement on the *Mesorhizobium* branch ($P \leqslant 0{\cdot}0165$). Disruption of the *Rhizobium leguminosarum–Rhizobium etli* lineage was significant in all cases ($P \leqslant 0{\cdot}0314$), confirming the neighbour-joining bootstrap support for this clade.

Evidence from insertions and deletions

There are a number of small insertions and deletions in the *atpD* alignment that provide additional phylogenetic evidence. Relative to all the other sequences (including outgroups), *Azorhizobium caulinodans* and *Brevundimonas bullata* have a 15 bp insertion at position 301 in the alignment and *Bradyrhizobium japonicum* has a 12 bp insertion at the same place. These insertions are related in sequence and provide good support for the monophyly of these three species, as shown in Fig. 2. All the sequences in Fig. 2, with the exception of those of *Brevundimonas bullata* and the two *Rhodobacter* species, have a 3 bp deletion at position 441 relative to the outgroups (*Rickettsia* has a 3 bp insertion at the same position). If the phylogeny in Fig. 2 is correct, this implies two independent deletions at the same site: one in the *Bradyrhizobium–Azorhizobium* lineage and the other in the ancestor of

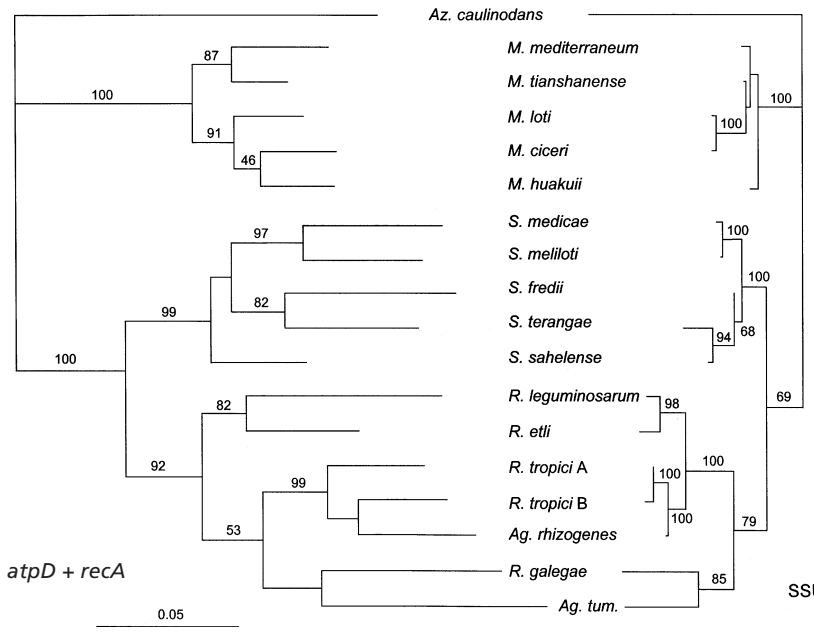

Fig. 5. Comparison of the combined *atpD–recA* phylogeny for rhizobia and agrobacteria and the corresponding SSU phylogeny. The trees were constructed by the maximum-likelihood method (F84 model), but bootstrap values (percentages of 1000 replicates) are for the corresponding branches in the neighbour-joining tree. Unlabelled branches were not found in the neighbour-joining tree. Genera are abbreviated as: *Ag.*, *Agrobacterium*; *Az.*, *Azorhizobium*; *M.*, *Mesorhizobium*; *R.*, *Rhizobium*; *S.*, *Sinorhizobium*.

the other species. A plausible alternative is that the *Rhodobacter* sequences are actually basal to the rest of the taxa in Fig. 2 and that there was a single deletion with a subsequent reinsertion in *Brevundimonas* (which has a high level of sequence divergence in general).

There are length variants in the SSU alignment, too, but they tend to occur in variable regions in which the exact alignment is uncertain. At around position 950, there are 3 bp missing in *Rhizobium leguminosarum*, *Rhizobium etli*, *Rhizobium tropici* A and B, *Rhizobium galegae* and *Agrobacterium rhizogenes*, but not in *Agrobacterium tumefaciens*. This unites the whole *Rhizobium–Agrobacterium* cluster except *Agrobacterium tumefaciens* and supports the bootscan analysis, which suggests that, in this segment of the SSU gene, *Agrobacterium tumefaciens* is divergent from *Rhizobium galegae* (see above). There is a similar 3 bp deletion in the *Brevundimonas bullata* sequence, but this presumably represents convergent evolution, since this sequence is very distant. A deletion of 21 bp at position 1200 is a shared derived character of *Brevundimonas* and the two *Rhodobacter* sequences, supporting a *Brevundimonas–Rhodobacter* clade that is not present in Fig. 1, but is not significantly contradicted. These three sequences also share a TTCG loop at positions 60–63 (all rhizobia have GCAA).

DISCUSSION

The *Mesorhizobium* clade

The genus *Mesorhizobium* was created to recognize a group of species with distinctive phenotypic properties and forming a clearly defined clade in the SSU phylogeny that was outside the *Rhizobium–Sinorhizobium–Agrobacterium* cluster (Jarvis *et al.*, 1997). Our analyses of *atpD* and *recA* sequences fully support both the integrity of the *Mesorhizobium* clade and its placement in the phylogeny. Further evidence is provided by a recent study of glutamine synthetase sequences (Turner & Young, 2000). The phylogeny of the GSI gene *glnA* yet again places *Mesorhizobium* as a tight clade outside the fast-growing rhizobia. However, the GSII gene is more equivocal. It is not possible to root the tree unambiguously because this gene is not widespread in bacteria, but the sequences for *Mesorhizobium mediterraneum*, *Mesorhizobium tianshanense*, *Mesorhizobium ciceri* and (less convincingly) *Mesorhizobium loti* do cluster together. The *Mesorhizobium huakuii* sequence, on the other hand, shows closer affinity with sequences of the genus *Rhizobium*. Turner & Young (2000) suggested that this might be the result of recombination; perhaps replacement of the whole gene, since they found no evidence for a recombination site within the gene. Our *atpD* and *recA* data reinforce the view that the SSU-based phylogeny holds for the majority of the genome and that GSII is exceptional and requires a special explanation of this kind.

Mesorhizobium forms a closely knit clade, but relationships among the species are not the same for all genes. For SSU, the *Mesorhizobium loti* and *Mesorhizobium ciceri* type strains are very similar, but this is not true for any of the other genes. This is consistent with the finding that *Mesorhizobium* strains that are virtually identical in SSU sequence may be unrelated in terms of DNA–DNA hybridization (Sullivan *et al.*, 1996) and suggests that there may be recombination of chromosomal genes between *Mesorhizobium* species.

The *Sinorhizobium* clade

Like *Mesorhizobium*, the genus *Sinorhizobium* forms a very clearly defined clade in the SSU tree, with a high level of sequence similarity among the species. All the

sequences from other genes confirm that *Sinorhizobium* is indeed a clade, with good bootstrap support: *atpD*, 94%; *recA*, 82% (*atpD* and *recA* jointly 99%), *glnA*, 96% and *glnII*, 100% (Turner & Young, 2000). All the genes also agree that *Sinorhizobium meliloti* and *Sinorhizobium medicae* are sister taxa, but *atpD*, *recA* and *glnA* all support a pairing of *Sinorhizobium terangae* with *Sinorhizobium fredii*, in contradiction to the pairing of *Sinorhizobium terangae* and *Sinorhizobium sahelense* that is strongly supported by SSU sequence and also favoured by *glnII*. Again, this might be interpreted as evidence for recombination between species.

Rhizobium and *Agrobacterium*

Almost 70 years ago, Fred *et al*. (1932) pointed out that agrobacteria and fast-growing rhizobia seemed to be more closely related than either were to bradyrhizobia, an insight that has of course been amply confirmed by modern methods. The amalgamation of *Rhizobium* and *Agrobacterium* has been suggested several times without a formal proposal (de Ley, 1968; Graham, 1964, 1976; Heberlein *et al*., 1967; Kerr, 1992; Sawada *et al*., 1993; White, 1972), but such a proposal has finally been made (Young *et al*., 2001). It is certainly true that the species of these two genera do not form two separate clades. One reason for this is easily dealt with: *Agrobacterium rhizogenes* (also known as biovar 2) is closely similar to *Rhizobium tropici* in SSU sequence and we see that this relationship also holds for *atpD* and *recA*. Clearly, these species belong in the same genus.

A deeper problem has been *Rhizobium galegae*, which usually clusters with *Agrobacterium* (specifically *Agrobacterium vitis*) rather than *Rhizobium* in the SSU phylogeny, as do *Rhizobium huautlense* (Wang *et al*., 1998) and *Allorhizobium undicola* (de Lajudie *et al*., 1998b). Wang *et al*. (1998) noted that, as more sequences were added to the *Rhizobium–Agrobacterium* part of the SSU phylogeny, the clustering of *Rhizobium galegae* with *Agrobacterium* disappeared. This is consistent with our evidence from the other genes: neither *atpD* nor *recA* provides any support for a *Rhizobium galegae–Agrobacterium tumefaciens* clade or for any other particular association involving either of these species.

Our data suggest that the distinction between *Rhizobium* and *Agrobacterium* is less clear-cut than it appears from SSU sequences alone, which could be taken as support for the proposal to amalgamate them (Young *et al*., 2001). Before doing this, though, it is important to be reasonably confident that the grouping is monophyletic. Considering only the species in our study, bootstrap support is 92% for SSU (Fig. 1) but only 48% for *recA* (Fig. 2), while in the *atpD* tree, *Mycoplana* and *Sinorhizobium* cannot be excluded from the clade. In the case of *recA*, the poor support is due largely to the two *Phyllobacterium* sequences, which are highly diverged but share a considerable number of phylogenetically informative sites with *Rhizobium–Agrobacterium*. The phylogenetic position of *Phyllobacterium* is uncertain, but the SSU and *atpD* phylogenies suggest that it does not belong in the *Rhizobium–Agrobacterium* clade. If only rhizobia and agrobacteria are considered, support for this clade rises to 93% (Fig. 5). Of course, there are several *Agrobacterium* and *Rhizobium* species that were not included in our study, and it will be important to examine their *atpD* and *recA* sequences before concluding that these genes fully support a *Rhizobium–Agrobacterium* clade.

Phyllobacterium, *Mycoplana* and *Brevundimonas*

While the close similarity of *Agrobacterium* to *Rhizobium* is indisputable, the phylogenetic placement of the other non-rhizobia is much more problematic. *Brevundimonas bullata* is always on a long branch with no close relatives, while the two *Mycoplana* species consistently cluster together, forming a separate clade at the same level as *Rhizobium*, *Sinorhizobium* and *Mesorhizobium* and not associated closely with any one of these genera. At all three loci, *Phyllobacterium myrsinacearum* and *Phyllobacterium rubiacearum* are so closely related that one might question whether they represent distinct species. Their phylogenetic position is, however, very uncertain. The SSU sequences have some affinity with *Mesorhizobium* and *recA* with *Rhizobium–Agrobacterium*, but the *Phyllobacterium* sequences are too diverged to place stably in the phylogenies. There is certainly little support for the suggestion that *Mesorhizobium* be placed together with *Phyllobacterium* in a family *Phyllobacteriaceae* (http://www.cme.msu.edu/Bergeys). There is little literature on *Phyllobacterium*, *Mycoplana* or *Brevundimonas* at present, but their pattern of molecular divergence suggests that their evolutionary history may be different from that of the rhizobia.

Next steps in the phylogeny of the rhizobia

Since this study was initiated, there have been several additions to the rhizobial species list as a result of sampling from a wider range of host-plant species and habitats. Some, such as *Rhizobium hainanense* (Chen *et al*., 1997), *Rhizobium gallicum* (Amarger *et al*., 1997), *Rhizobium mongolense* (van Berkum *et al*., 1998), *Mesorhizobium amorphae* (Wang *et al*., 1999), *Mesorhizobium plurifarium* (de Lajudie *et al*., 1998a), *Sinorhizobium arboris* and *Sinorhizobium kostiense* (Nick *et al*., 1999), can be placed very firmly within existing clades on the basis of their SSU sequences. While it would be reassuring to confirm these placements using other genes, we would not expect many surprises because our data show that an unambiguous SSU similarity is usually reflected in other genes. More interesting are the new species of less obvious affinity, and further gene sequences from *Rhizobium giardinii*, *Rhizobium huautlense* and *Allorhizobium undicola* would be welcome.

Despite a greatly broadened sampling of host-plant species and geographical locations, the taxonomic diversity of root-nodule bacteria has, until recently, remained relatively narrow and clearly defined. The new species have fallen mainly within the existing genera or, failing that, between them rather than beyond them. The recent report (Sy *et al*., 2001) of root-nodule bacteria in the genus *Methylobacterium* is an exception and represents the first expansion of the known phylogenetic range of rhizobia since the description of *Azorhizobium* (Dreyfus *et al*., 1988). Nevertheless, plasmid replicons related to those that carry symbiosis genes in many rhizobia have been found in *Paracoccus* (Bartosik *et al*., 1997) and *Rhodobacter* (unpublished *Rhodobacter capsulatus* genome sequence at http://rhodo.img.cas.cz/) and there seems no intrinsic reason why the root-nodule symbiosis should not extend to these groups of the α-*Proteobacteria* or even beyond. The broad phylogenetic agreement that we report here is true only for the 'housekeeping' genome and there is now ample evidence that the symbiosis genes have a quite different evolutionary history as a result of horizontal transfer. There are probably other sets of 'contingency genes' that have idiosyncratic phylogenies, but our study indicates that there is a robust and consistent phylogeny for the genomic backbone of housekeeping genes in bacteria.

ACKNOWLEDGEMENTS

We are grateful for conversations with Sandie Baldauf and Gary Olsen. We thank Peter van Berkum, Kristina Lindström, Sophie Rome and David Stead for sending strains, Allen Mould for DNA sequencing, Jeremy Searle for access to PAUP* and John Robinson and staff at both York and Manchester computer services for technical advice and support. A large proportion of the computation was conducted on the CSAR Origin2000 machine at Manchester Computing. The work was funded by NERC grant GR3/10819 and contract BIO4-CT98-0483 (TRAFFIC) from the Commission of the European Communities.

REFERENCES

Abraham, W.-R., Strömpl, C., Meyer, H. & 8 other authors (1999). Phylogeny and polyphasic taxonomy of *Caulobacter* species. Proposal of *Maricaulis* gen. nov. with *Maricaulis maris* (Poindexter) comb. nov. as the type species, and emended description of the genera *Brevundimonas* and *Caulobacter*. *Int J Syst Bacteriol* **49**, 1053–1073.

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

Andersson, S. G. E., Zomorodipour, A., Andersson, J. O. & 7 other authors (1998). The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* **396**, 133–140.

Barrera, L. L., Trujillo, M. E., Goodfellow, M., García, F. J., Hernández-Lucas, I., Dávila, G., van Berkum, P. & Martínez-Romero, E. (1997). Biodiversity of bradyrhizobia nodulating *Lupinus* spp. *Int J Syst Bacteriol* **47**, 1086–1091.

Bartosik, D., Wlodarczyk, M. & Thomas, C. M. (1997). Complete nucleotide sequence of the replicator region of *Paracoccus* (*Thiobacillus*) *versutus* pTAV1 plasmid and its correlation to several plasmids of *Agrobacterium* and *Rhizobium* species. *Plasmid* **38**, 53–59.

Beringer, J. E. (1974). R factor transfer in *Rhizobium leguminosarum*. *J Gen Microbiol* **84**, 188–198.

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

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

Borghese, R., Crimi, M., Fava, L. & Melandri, B. A. (1998). The ATP synthase atpHAGDC (F1) operon from *Rhodobacter capsulatus*. *J Bacteriol* **180**, 416–421.

Calero, S., Fernandez de Henestrosa, A. R. & Barbe, J. (1994). Molecular cloning, sequence and regulation of expression of the recA gene of the phototrophic bacterium *Rhodobacter sphaeroides*. *Mol Gen Genet* **242**, 116–120.

Chen, W.-X., Tan, Z.-Y., Gao, J.-L., Li, Y. & Wang, E.-T. (1997). *Rhizobium hainanense* sp. nov., isolated from tropical legumes. *Int J Syst Bacteriol* **47**, 870–873.

Cunningham, C. W. (1997). Can three incongruence tests predict when data should be combined? *Mol Biol Evol* **14**, 733–740.

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

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

Dryden, S. C. & Kaplan, S. (1990). Localization and structural analysis of the ribosomal RNA operons of *Rhodobacter sphaeroides*. *Nucleic Acids Res* **18**, 7267–7277.

Dunkin, S. M. & Wood, D. O. (1994). Isolation and characterization of the *Rickettsia prowazekii* recA gene. *J Bacteriol* **176**, 1777–1781.

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

Eckloff, B. W., Podzorski, R. P., Kline, B. C. & Cockerill, F. R., III (1994). A comparison of 16S ribosomal DNA sequences from five isolates of *Helicobacter pylori*. *Int J Syst Bacteriol* **44**, 320–323.

Eisen, J. A. (1995). The RecA protein as a model molecule for molecular systematic studies of bacteria: comparison of trees of RecAs and 16S rRNAs from the same species. *J Mol Evol* **41**, 1105–1123.

Farris, J. S., Källersjö, M., Kluge, A. G. & Bult, C. (1994). Testing significance of incongruence. *Cladistics* **10**, 315–319.

Fernandez de Henestrosa, A. R., Rivera, E. & Barbe, J. (1995). Non-reciprocal regulation of *Rhodobacter capsulatus* and *Rhodobacter sphaeroides* recA genes expression. *FEMS Microbiol Lett* **129**, 175–181.

Fred, E. B., Baldwin, I. L. & McCoy, E. (1932). *Root Nodule Bacteria and Leguminous Plants*. Madison, WI: University of Wisconsin.

Galtier, N., Gouy, M. & Gautier, C. (1996). SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. *Comput Appl Biosci* **12**, 543–548.

Graham, P. H. (1964). The application of computer techniques to the taxonomy of the root-nodule bacteria of legumes. *J Gen Microbiol* **35**, 511–517.

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

Heberlein, G. T., De Ley, J. & Tijtgat, R. (1967). Deoxyribonucleic acid homology and taxonomy of *Agrobacterium*, *Rhizobium*, and *Chromobacterium*. *J Bacteriol* **94**, 116–124.

Hillis, D. M. & Huelsenbeck, J. P. (1992). Signal, noise, and reliability in molecular phylogenetic analyses. *J Hered* **83**, 189–195.

Hiraishi, A. & Ueda, Y. (1994). Intrageneric structure of the genus *Rhodobacter*: transfer of *Rhodobacter sulfidophilus* and related marine species to the genus *Rhodovulum* gen. nov. *Int J Syst Bacteriol* **44**, 15–23.

Jarvis, B. D. W., Van Berkum, P., Chen, W. X., Nour, S. M., Fernandez, M. P., Cleyet-Marel, J. C. & Gillis, M. (1997). Transfer of *Rhizobium loti*, *Rhizobium huakuii*, *Rhizobium ciceri*, *Rhizobium mediterraneum*, and *Rhizobium tianshanense* to *Mesorhizobium* gen. nov. *Int J Syst Bacteriol* **47**, 895–898.

Jordan, D. C. (1982). Transfer of *Rhizobium japonicum* Buchanan 1980 to *Bradyrhizobium* gen. nov., a genus of slow-growing, root nodule bacteria from leguminous plants. *Int J Syst Bacteriol* **32**, 136–139.

Kaijalainen, S. & Lindström, K. (1989). Restriction fragment length polymorphism analysis of *Rhizobium galegae* strains. *J Bacteriol* **171**, 5561–5566.

Kaneko, T., Nakamura, Y., Sato, S. & 21 other authors (2000). Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*. *DNA Res* **7**, 331–338.

Kerr, A. (1992). The genus *Agrobacterium*. In *The Prokaryotes – a Handbook on the Biology of Bacteria*, pp. 2214–2235. Edited by A. Balows, H. G. Trüper, M. Dworkin, W. Harder & K.-H. Schleifer. Heidelberg: Springer.

Laguerre, G., Mazurier, S. I. & Amarger, N. (1992). Plasmid profiles and restriction fragment length polymorphism of *Rhizobium leguminosarum* bv. *viciae* in field populations. *FEMS Microbiol Lett* **101**, 17–26.

de Lajudie, P., Willems, A., Pot, B. & 7 other authors (1994). Polyphasic taxonomy of rhizobia: emendation of the genus *Sinorhizobium* and description of *Sinorhizobium meliloti* comb. nov., *Sinorhizobium saheli* sp. nov., and *Sinorhizobium teranga* sp. nov. *Int J Syst Bacteriol* **44**, 715–733.

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

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

Lewis, P. O. (1998). A genetic algorithm for maximum-likelihood phylogeny inference using nucleotide sequence data. *Mol Biol Evol* **15**, 277–283.

de Ley, J. (1968). DNA base composition and hybridization in the taxonomy of the phytopathogenic bacteria. *Annu Rev Phytopathol* **6**, 63–90.

Liesack, W., Soller, R., Stewart, T., Haas, H., Giovannoni, S. & Stackebrandt, E. (1992). The influence of tachytelically (rapidly) evolving sequences on the topology of phylogenetic trees: intrafamily relationships and the phylogenetic position of *Planctomycetaceae* as revealed by comparative analysis of 16S ribosomal RNA sequences. *Syst Appl Microbiol* **15**, 357–362.

Lindström, K., Paulin, L., Roos, C. & Suominen, L. (1995). Nodulation genes of *Rhizobium galegae*. In *Nitrogen Fixation: Fundamentals and Applications*, pp. 365–370. Edited by I. A. Tikhonovich, N. A. Provorov, V. I. Romanov & W. E. Newton. Dordrecht: Kluwer.

McGowan, C. C., Cover, T. L. & Blaser, M. J. (1997). Analysis of F_1F_0-ATPase from *Helicobacter pylori*. *Infect Immun* **65**, 2640–2647.

Maddison, W. P. & Maddison, D. R. (1992). *MacClade: Analysis of Phylogeny and Character Evolution*. Sunderland, MA: Sinauer.

Maidak, B. L., Cole, J. R., Lilburn, T. G. & 9 other authors (2000). The RDP (Ribosomal Database Project) continues. *Nucleic Acids Res* **28**, 173–174.

Mickevich, M. F. & Farris, J. S. (1981). The implications of congruence in *Menidia*. *Syst Zool* **30**, 351–370.

Moreno, E., Stackebrandt, E., Dorsch, M., Wolters, J., Busch, M. & Mayer, H. (1990). *Brucella abortus* 16S rRNA and lipid A reveal a phylogenetic relationship with members of the alpha-2 subdivision of the class *Proteobacteria*. *J Bacteriol* **172**, 3569–3576.

Nick, G., de Lajudie, P., Eardly, B. D., Suomalainen, S., Paulin, L., Zhang, X., Gillis, M. & Lindström, K. (1999). *Sinorhizobium arboris* sp. nov. and *Sinorhizobium kostiense* sp. nov., isolated from leguminous trees in Sudan and Kenya. *Int J Syst Bacteriol* **49**, 1359–1368.

Nour, S. M., Fernandez, M. P., Normand, P. & Cleyet-Marel, J.-C. (1994). *Rhizobium ciceri* sp. nov., consisting of strains that nodulate chickpeas (*Cicer arietinum* L.). *Int J Syst Bacteriol* **44**, 511–522.

Nour, S. M., Cleyet-Marel, J.-C., Normand, P. & Fernandez, M. P. (1995). Genomic heterogeneity of strains nodulating chickpeas (*Cicer arietinum* L.) and description of *Rhizobium mediterraneum* sp. nov. *Int J Syst Bacteriol* **45**, 640–648.

Olsen, G. J., Matsuda, H., Hagstrom, R. & Overbeek, R. (1994). fastDNAmL: a tool for construction of phylogenetic trees of DNA sequences using maximum likelihood. *Comput Appl Biosci* **10**, 41–48.

Pamilo, P. & Nei, M. (1988). Relationships between gene trees and species trees. *Mol Biol Evol* **5**, 568–583.

Pueppke, S. G. (1996). The genetic and biochemical basis for nodulation of legumes by rhizobia. *Crit Rev Biotechnol* **16**, 1–51.

Rome, S., Fernandez, M. P., Brunel, B., Normand, P. & Cleyet-Marel, J.-C. (1996). *Sinorhizobium medicae* sp. nov., isolated from annual *Medicago* spp. *Int J Syst Bacteriol* **46**, 972–980.

Ronner, S., Liesack, W., Wolters, J. & Stackebrandt, E. (1991). Cloning and sequencing of a large fragment of the atpD-gene of *Pirellula marine*: a contribution to the phylogeny of planctomycetales. *Endocytobiosis Cell Res* **7**, 219–229.

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

Sakurada, K., Ohta, T., Fujishiro, K., Hasegawa, M. & Aisaka, K. (1996). Acetylpolyamine amidohydrolase from *Mycoplana ramosa*: gene cloning and characterization of the metal-substituted enzyme. *J Bacteriol* **178**, 5781–5786.

Salminen, M. O., Carr, J. K., Burke, D. S. & McCutchan, F. E.

(1995). Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning. *AIDS Res Hum Retroviruses* **11**, 1423–1425.

Sawada, H., Ieki, H., Oyaizu, H. & Matsumoto, S. (1993). Proposal for rejection of *Agrobacterium tumefaciens* and revised descriptions for the genus *Agrobacterium* and for *Agrobacterium radiobacter* and *Agrobacterium rhizogenes*. *Int J Syst Bacteriol* **43**, 694–702.

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.

Swofford, D. L. (1997). *Phylogenetic Analysis using Parsimony – PAUP**. Sunderland, MA: Sinauer.

Sy, A., Giraud, E., Jourand, P. & 8 other authors (2001). Methylotrophic *Methylobacterium* bacteria nodulate and fix nitrogen in symbiosis with legumes. *J Bacteriol* **183**, 214–220.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**, 4876–4882.

Thompson, S. A. & Blaser, M. J. (1995). Isolation of the *Helicobacter pylori* recA gene and involvement of the recA region in resistance to low pH. *Infect Immun* **63**, 2185–2193.

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

Tybulewicz, V. L. J., Falk, G. & Walker, J. E. (1984). *Rhodopseudomonas blastica atp* operon: nucleotide sequence and transcription. *J Mol Biol* **179**, 185–214.

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

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

Weisburg, W. G., Woese, C. R., Dobson, M. E. & Weiss, E. (1985). A common origin of rickettsiae and certain plant pathogens. *Science* **230**, 556–558.

Weisburg, W. G., Dobson, M. E., Samuel, J. E. & 7 other authors (1989). Phylogenetic diversity of the rickettsiae. *J Bacteriol* **171**, 4202–4206.

Wheatcroft, R., McRae, D. G. & Miller, R. W. (1990). Changes in the *Rhizobium meliloti* genome and the ability to detect supercoiled plasmids during bacteroid development. *Mol Plant Microbe Interact* **3**, 9–17.

White, L. O. (1972). The taxonomy of the crown-gall organism *Agrobacterium tumefaciens* and its relationship to rhizobia and other agrobacteria. *J Gen Microbiol* **72**, 565–574.

Willems, A. & Collins, M. D. (1993). Phylogenetic analysis of rhizobia and agrobacteria based on 16S rRNA gene sequences. *Int J Syst Bacteriol* **43**, 305–313.

Yanagi, M. & Yamasato, K. (1993). Phylogenetic analysis of the family *Rhizobiaceae* and related bacteria by sequencing of 16S rRNA gene using PCR and DNA sequencer. *FEMS Microbiol Lett* **107**, 115–120.

Yang, Z., Goldman, N. & Friday, A. (1994). Comparison of models for nucleotide substitution used in maximum-likelihood phylogenetic estimation. *Mol Biol Evol* **11**, 316–324.

Young, J. M., Kuykendall, L. D., Martínez-Romero, E., Kerr, A. & Sawada, H. (2001). 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**, 89–103.

Young, J. P. W. (1996). Phylogeny and taxonomy of rhizobia. *Plant Soil* **186**, 45–52.

Young, J. P. W. (1998). Bacterial evolution and the nature of species. In *Advances in Molecular Ecology*, pp. 119–131. Edited by G. R. Carvalho. Amsterdam: IOS Press.

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

Young, J. P. W. & Johnston, A. W. B. (1989). The evolution of specificity in the legume-rhizobium symbiosis. *Trends Ecol Evol* **4**, 341–349.

Young, J. P. W., Downer, H. L. & Eardly, B. D. (1991). Phylogeny of the phototrophic rhizobium strain BTAi1 by polymerase chain reaction-based sequencing of a 16S rRNA gene segment. *J Bacteriol* **173**, 2271–2277.