Molecular Phylogenetics and Evolution 48 (2008) 1131–1144

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Cowpea and peanut in southern Africa are nodulated by diverse *Bradyrhizobium* strains harboring nodulation genes that belong to the large pantropical clade common in Africa

Emma T. Steenkamp [a,*], Tomasz Stępkowski [b], Anna Przymusiak [b], Wilhelm J. Botha [c], Ian J. Law [c]

[a] Department of Microbiology and Plant Pathology, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Pretoria 0002, South Africa
[b] Institute of Bioorganic Chemistry, Polish Academy of Sciences, 61-704 Poznan, Noskowskiego 12/14, Poland
[c] Agricultural Research Council (ARC), Plant Protection Research Institute, Private Bag X134, Queenswood 0121, South Africa

ARTICLE INFO

Article history:
Received 23 January 2008
Revised 13 April 2008
Accepted 23 April 2008
Available online 1 May 2008

Keywords:
Bradyrhizobium
Biogeography
Vigna unguiculata
nodA
noeE
glnII
recA
rrs
Horizontal gene transfer

ABSTRACT

Cowpea (*Vigna unguiculata*) and peanut (*Arachis hypogaea*) in southern Africa are nodulated by a genetically diverse group of *Bradyrhizobium* strains. To determine the identity of these bacteria, a collection of 22 isolates originating from the root nodules of both hosts in Botswana and South Africa was investigated using the combined sequences for the core genome genes *rrs*, *recA*, and *glnII*. These data separated the majority of the isolates into one of three unique lineages that most likely represent novel *Bradyrhizobium* species. Some isolates were also conspecific with *B. yuanmingense* and with *B. elkanii*, although none grouped with *B. japonicum*, *B. canariense* or *B. liaoningense*. To study the evolution of nodulation genes in these bacteria, the common nodulation gene, *nodA*, and host-specific nodulation genes, *nodZ*, *noeE*, and *noeI*, were analyzed. The *nodA* phylogeny showed that the cowpea and peanut *Bradyrhizobium* isolates represent various locally adapted groups or ecotypes that form part of Clade III of the seven known *Bradyrhizobium nodA* clades. This large and highly diverse clade comprises all strains from sub-Saharan Africa, as well as some originating from the Americas, Australia, Indonesia, China and Japan. Some similar groupings were supported by the other nodulation genes, although the overall phylogenies for the nodulation genes were incongruent with that inferred from the core genome genes, suggesting that horizontal gene transfer significantly influences the evolution of cowpea and peanut root-nodule bacteria. Furthermore, identification of the *nodZ*, *noeI*, and *noeE* genes in the isolates tested indicates that African *Bradyrhizobium* species may produce highly decorated nodulation factors, which potentially represent an important adaptation enabling nodulation of a great variety of legumes inhabiting the African continent.
© 2008 Elsevier Inc. All rights reserved.

1. Introduction

Cowpea, *Vigna unguiculata* (L.) Walpers, a legume native to Africa, is an important annual crop in tropical and sub-tropical regions worldwide, especially in sub-Saharan Africa, Asia, and Central and South America (Duke, 1981; Singh et al., 2003; Steele et al., 1985). The young leafs, pods, and seeds of this plant are good sources of dietary protein, vitamins, and minerals for humans and animals (Barrett, 1990; Bressani, 1985; Singh et al., 2003). Cultivation of cowpea and other legumes also help maintain soil quality through biological nitrogen fixation by symbiotic rhizobial bacteria (Dakora and Keya, 1997; Graham and Vance, 2003). Many soils in cultivated areas contain rhizobia from the so-called 'cowpea group' that effectively nodulate this and other tropical hosts, yet agricultural practices usually include application of commercial inoculant strains (Allen and Allen, 1981; Singleton et al., 1991; Strijdom, 1998).

Rhizobia representing the 'cowpea group' are usually slow-growing bacteria of the genus *Bradyrhizobium* in the order Rhizobiales of the class Alphaproteobacteria (Garrity et al., 2005a; Jordan, 1989; Kuykendall, 2005a; Kuykendall, 2005b). *Bradyrhizobium* is a cosmopolitan and diverse group capable of nodulating a variety of legumes, as well as the non-legume host *Parasponia* (Kuykendall, 2005a; Lafay et al., 2006). A member of this genus has also been associated with tumour-like deformations on the roots of sugar beet (Rivas et al., 2004). Despite their wide-ranging and overlapping metabolic abilities and lifestyles (Kuykendall, 2005a), *Bradyrhizobium* species are readily distinguishable using DNA sequence information for marker-genes that form part of the so-called "core genome". For example, sequence analyses of the internally transcribed spacer (ITS) region between the 16S (*rrs*) and 23S (*rrl*) ribosomal RNA (rRNA) genes and the genes encoding recombination protein RecA (*recA*) and glutamine synthetase isoform II (*glnII*), allow separation of this genus into seven described species (*B. japonicum*, *B. elkanii*, *B. liaoningense*, *B. yuanmingense*, *B. canariense*, *B. betae*, and *B. denitrificans*), two so-called genospecies and a

* Corresponding author. Fax: +27 12 420 3266.
E-mail address: emma.steenkamp@fabi.up.ac.za (E.T. Steenkamp).

1055-7903/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2008.04.032

number of distinct phylogenetic lineages (Rivas et al., 2004; Stępkowski et al., 2007, 2005; van Berkum et al., 2006; Vinuesa et al., 2005a,b; Weir et al., 2004).

At the genetic level, rhizobial nodulation is controlled by the nodulation genes *nod*, *noe*, and *nol* (Dénarié et al., 1996; Schultze and Kondorosi, 1998). These genes form part of the accessory genome *sensu* Young et al. (2006) and appear to be dispensable for cell function (Stępkowski and Legocki, 2001). In the model *Bradyrhizobium* species, *B. japonicum*, the nodulation genes, together with those controlling nitrogen fixation (*nif* genes), are flanked by insertion sequence (IS) elements in a large symbiosis (*sym*) island that is integrated into a valine tRNA-encoding gene (Kaluza et al., 1985; Kaneko et al., 2002). The nodulation genes determine the formation and structure of Nod factors (NFs), which represent the primary signaling molecules implicated in host infection and nodule organogenesis (van Rhijn and Vanderleyden, 1995). The structural *nodABC* genes are absolutely essential for symbiosis and their products (NodA, acyl transferase; NodB, deacetylase; NodC, *N*-acetylglucosaminyl transferase) produce a β-1,4-linked *N*-acylated-D-glucosamine backbone molecule consisting of three to five sugar units and a fatty-acid group (with varied length and saturation) attached to the non-reducing end. This lipo-chitooligosaccharide constitutes the NF core and may be decorated by the products of a class of nodulation genes that are often referred to as host-specific nodulation (*hsn*) genes. Various studies have shown that such modifications of the NF core, especially at the reducing end, are important determinants of the host range of a specific rhizobial strain (e.g., Carlson et al., 1993; Lopez-Lara et al., 1996; Perret et al., 2000). As rhizobia use different sets of *hsn* genes for decorating their NFs, rhizobial species and lineages differ broadly in their complement of these genes (e.g., Moulin et al., 2004; van Rhijn and Vanderleyden, 1995).

Numerous studies have demonstrated that horizontal gene transfer (HGT) plays a significant role in the evolution of rhizobia (e.g., Flores et al., 2005; Sullivan and Ronson, 1998; Young et al., 2006), as well as their host range (e.g., Moulin et al., 2004; Suominen et al., 2001; Vinuesa et al., 2005b). In some instances, HGT may result in "patchy" compositions of individual genomes where not all individuals have the same complement of genes (Koonin et al., 2001; Ochman et al., 2000; Welch et al., 2002) as is evident from the erratic distribution of nodulation genes among rhizobial species and lineages (Moulin et al., 2004). Also, the evolutionary histories of genes acquired through HGT (and usually genes encoded on the accessory genome) are mostly incongruent with those of genes that form part of the core genome (Koonin et al., 2001; Philippe and Douady, 2003). Phylogenetic analyses of nodulation or *nif* genes typically reveal substructures in rhizobial populations that are remarkably different from those using core genome genes (e.g., Laguerre et al., 2001; Mutch et al., 2003; Qian and Parker, 2002). For example, *Bradyrhizobium nodC* and *nodA* sequences from a specific host and/or geographic region, regardless of the *Bradyrhizobium* species harboring the sequence, are more closely related to one another than to those from other legume hosts and/or geographic regions (Stępkowski et al., 2005; Vinuesa et al., 2005b). Current models for the evolution of rhizobia therefore hold that different rhizobial lineages in various geographic areas, diversify through HGT of *sym* loci to generate locally adapted genotypes or so-called biovarieties that are associated with a specific group of hosts (Laguerre et al., 2001; Mutch and Young, 2004; Parker et al., 2002; Vinuesa et al., 2005a,b).

A recent study demonstrated that *Bradyrhizobium* isolates associated with the root nodules of cowpea and peanut (*Arachis hypogaea*) in Botswana and South Africa are genetically diverse (Law et al., 2007). Genomic fingerprints of these bacteria revealed several clusters of isolates that are less than 60% similar. This diversity was also reflected by the symbiotic properties of these bacteria, as they displayed a range of nodulation and nitrogen fixation abilities on one or both hosts. The aim of the present study was to determine the extent to which the observed diversity in this collection of 'cowpea group' *Bradyrhizobium* isolates may be ascribed to the processes of vertical descent due to divergence of the core genome and subsequent inheritance in progeny and to acquisition of specific nodulation genes via HGT. For this purpose we first selected a diverse range of *Bradyrhizobium* isolates from the larger set of strains obtained from different localities in South Africa and Botswana, using ITS PCR-RFLP (restriction fragment length polymorphism) analyses. Second, we determined the gene sequences for *glnII*, *recA*, and *rrs* located on the core genome, as well as *nodA*, *nodZ*, *noeE*, and *noeI* encoded at the *sym* loci in the selected *Bradyrhizobium* isolates. Phylogenetic analyses of the core genome regions were used to determine the possible identities of the bacteria, while comparisons of core and nodulation gene trees were used to detect and study HGT.

2. Materials and methods

2.1. Bacterial isolates

The location, isolation, and strain diversity of 74 'cowpea group' *Bradyrhizobium* isolates from soils in Gaborone, Francistown, Good Hope, Maun, and Rasesa in Botswana and Roodeplaat in South Africa was described in an earlier study (Law et al., 2007). None of the soils included in that study had a history of inoculation with rhizobial strains (Law et al., 2007). Twenty-two *Bradyrhizobium* isolates (of which five were isolated from peanut root nodules) that represent the thirteen diversity groups identified by ITS PCR-RFLP in the present study (see below), were selected for sequence analysis (Table 1). We also included the cowpea inoculant strain CB756, as well as the *B. japonicum* (LMG6138) and *B. elkanii* (LMG6134) type strains. All isolates are maintained in the South African Rhizobium Collection (Agricultural Research Council, Plant Protection Research Institute, Roodeplaat, South Africa). Yeast extract mannitol medium (Somasegaran and Hoben, 1994) and incubation at 25–28°C were used for routine growth of the isolates.

2.2. ITS PCR-RFLP analysis

Whole-cell DNA was extracted as described before by Botha et al. (2002). The rRNA ITS region for the 74 'cowpea group' *Bradyrhizobium* isolates was amplified with primer set SM + BR3 (Willems et al., 2001a), using Promega PCR buffer and *Taq* polymerase (Promega Corp., Madison, WI, USA) and previously described reaction and cycling conditions (Willems et al., 2001a). Each of the resulting amplicons were then digested at 37 °C for 4 h in the presence of 0.2 U/μl of the restriction endonucleases AluI and HaeIII (Promega) and commercially supplied digestion buffer. Restriction products were separated with agarose (2%, w/v) gel electrophoresis and results were recorded using previously described information capture methods (Botha et al., 2004). Bionumerics version 4.0 (Applied-Maths, Sint-Martens-Latem, Belgium) was used to normalize gel images and to convert RFLP profiles into binary matrices based on the presence and absence of restriction fragments. This software was also used to construct UPGMA dendograms from the binary matrices using the Dice similarity coefficient.

2.3. Amplification and sequencing of the rrs, recA, and glnII genes

In this study we utilized the sequence information for *rrs*, *glnII*, and *recA*, as extensive nucleotide information for the known species and lineages of *Bradyrhizobium* are available in public domain

Table 1
Origin, identity, cluster and sequence information for the *Bradyrhizobium* isolates used in this study

Isolate	Geographic origin	Host	*rrs* + *recA* + *glnII* lineage[a]	*nodA* Clade[b]	ITS PCR-RFLP diversity group[c]	GenBank Accession Nos.[d]
BM1	Maun, Botswana	*V. unguiculata*	SA-3	III.3-D	7	EU364697(S), EU364674(R), EU364651(G), EU364720(A)
BM25	Maun, Botswana	*V. unguiculata*	SA-3	III.3-D	7	EU364698(S), EU364675(R), EU364652(G), EU364721(A), EU364743(Z), EU364752(E)
FC1b	Francistown, Botswana	*V. unguiculata*	SA-3	III.3-A	2	EU364700(S), EU364677(R), EU364654(G), EU364723(A), EU364744(Z), EU364754(E), EU364762(I)
FP1c	Francistown, Botswana	*A. hypogaea*	SA-3	III.3-A	2	EU364701(S), EU364678(R), EU364655(G), EU364724(A)
FP4f	Francistown, Botswana	*A. hypogaea*	SA-3	III.3-A	1	EU364702(S), EU364679(R), EU364656(G), EU364725(A)
GC1d	Gaborone, Botswana	*V. unguiculata*	SA-1	III.3-C	12	EU364703(S), EU364680(R), EU364657(G), EU364726(A), EU364749(Z), EU364755(E), EU364763(I)
GP2e	Gaborone, Botswana	*A. hypogaea*	SA-3	III.3-A	6	EU364708(S), EU364685(R), EU364662(G), EU364731(A)
GHa	Good Hope, Botswana	*V. unguiculata*	SA-1	III.3-G	11	EU364704(S), EU364681(R), EU364658(G), EU364727(A), EU364750(Z), EU364756(E), EU364764(I)
GHx	Good Hope, Botswana	*V. unguiculata*	SA-1	III.3-G	11	EU364707(S), EU364684(R), EU364661(G), EU364730(A)
GHiv	Good Hope, Botswana	*V. unguiculata*	SA-1	III.3-C	12	EU364705(S), EU3646820(R), EU364659(G), EU364728(A)
GHvi	Good Hope, Botswana	*V. unguiculata*	SA-1	III.3-C	10	EU364706(S), EU364683(R), EU364660(G), EU364729(A)
R2m	Rasesa, Botswana	*V. unguiculata*	*B. yuanmingense*	III.3-F	4	EU364709(S), EU364686(R), EU364663(G), EU364732(A), EU364747(Z), EU364757(E), EU364765(I)
R3	Rasesa, Botswana	*V. unguiculata*	*B. yuanmingense*	III.3-B	3	EU364710(S), EU364687(R), EU364664(G), EU364733(A), EU364745(Z), EU364758(E), EU364766(I)
R5	Rasesa, Botswana	*V. unguiculata*	*B. elkanii*	III.3-H	9	EU364711(S), EU364688(R), EU364665(G), EU364734(A), EU364746(Z), EU364767(I)
R8	Rasesa, Botswana	*V. unguiculata*	*B. yuanmingense*	III.3-A	5	EU364712(S), EU364689(R), EU364666(G), EU364735(A)
R8m	Rasesa, Botswana	*V. unguiculata*	*B. yuanmingense*	III.3-F	13	EU364713(S), EU364690(R), EU364667(G), EU364736(A)
R10	Rasesa, Botswana	*V. unguiculata*	*B. yuanmingense*	III.3-F	13	EU364714(S), EU364691(R), EU364668(G), EU364737(A)
R10m	Rasesa, Botswana	*V. unguiculata*	*B. yuanmingense*	III.3-F	3	EU364715(S), EU364692(R), EU364669(G), EU364738(A)
RC2d	Roodeplaat, South Africa	*V. unguiculata*	SA-3	III.3-A	6	EU364716(S), EU364693(R), EU364670(G), EU364739(A)
RC3b	Roodeplaat, South Africa	*V. unguiculata*	SA-2	III.3-E	8	EU364717(S), EU364694(R), EU364671(G), EU364740(A), EU364759(E), EU364768(I)
RP6b	Roodeplaat, South Africa	*A. hypogaea*	SA-3	III.3-A	5	EU364718(S), EU364695(R), EU364672(G), EU364741(A), EU364751(Z), EU364760(E), EU364769(I)
RP7b	Roodeplaat, South Africa	*A. hypogaea*	SA-3	III.3-A	1	EU364719(S), EU364696(R), EU3646673(G), EU364742(A)
CB756	Zimbabwe	*Macrotyloma africanum*	SA-4	III.3-E	3	EU364699(S), EU364676(R), EU364653(G), EU364722(A), EU364748(Z), EU364753(E), EU364761(I)
LMG6138	Japan	*Glycine max*	*B. japonicum*			
LMG6134	United States of America	*G. max*	*B. elkanii*			

[a, b, and c] See Figs. 2, 3 and 1, respectively.
[d] S, R, G, A, Z, E, and I in parentheses refer to *rrs*, *recA*, *glnII*, *nodA*, *nodZ*, *noeE*, and *noeI* sequences, respectively.

nucleotide databases. Total DNA was isolated from the selected 22 'cowpea group' isolates, as well as the inoculant and reference isolates of *Bradyrhizobium* using the DNeasy Tissue kit (QIAGEN, Germany). For all these isolates, 1500 base pair (bp), 600 and 500 bp portions of the genes encoding *rrs*, *recA*, and *glnII* were PCR-amplified using the primer sets 16S-F + 16S-R (Molouba et al., 1999), TSrecAf + TSrecAr (Stępkowski et al., 2005) and *glnII* 12F + *glnII* 689R (Vinuesa et al., 2005b), respectively. Amplification of these regions was accomplished using FastStart *Taq* DNA polymerase (Roche Diagnostics GmbH, Germany) and previously described reaction and cycling conditions (Molouba et al., 1999; Stępkowski et al., 2005; Vinuesa et al., 2005b). PCR products were purified by polyethylene glycol precipitation (Steenkamp et al., 2006) and subjected to automated sequencing using the original PCR primers and Applied Biosystems' ABI PRISM BigDye Terminator v3.0 Cycle Sequencing Kit and the 3730 DNA Analyzer (Applied Biosystems, Foster City, Calif.).

2.4. Amplification and sequencing of the nodA, nodZ, noeE, and noeI genes

Total genomic DNA was isolated from the 22 *Bradyrhizobium* isolates identified in the ITS PCR-RFLP analysis, as well as CB756 and the two type strains. For this purpose, the QIAamp DNA mini kit (QIAGEN) or the sodium dodecyl sulfate–proteinase K lysis procedure (Moulin et al., 2004) was used. For all of the isolates, a fragment comprising the intergenic region between *nodD* and *nodA*, the entire *nodA* gene, and part of the *nodB* gene, was amplified using the primers TSnodD1-1c and TSnodB2N (Stępkowski et al., 2005). From selected isolates, portions of the genes encoding *nodZ* and *noeI* were amplified with the primer sets TSnodZ3 + TSnodZ4, and TSnoeI1 + TSnoeI2 (Moulin et al., 2004). These amplification reactions utilized FastStart *Taq* DNA polymerase and previously described reaction and cycling conditions (Moulin et al., 2004; Stępkowski et al., 2005).

We also amplified a portion of the *noeE* gene using primers TSnoeE1 (GAGGCGTTTGGCARGGTGGA) and TSnoeE3 (CGGTYAYSGTTTGCATCTCYTCGA). These primers were designed using the single *noeE* gene sequence available (Freiberg et al., 1997) together with the interrupted *noeE* gene sequences from soybean strains, *B. japonicum* USDA110 (Kaneko et al., 2002) and *B. elkanii* USDA94 (Yasuta et al., 2001). They amplify a region corresponding to nucleotide positions 142–817 of the *noeE* gene of *Sinorhizobium* sp. NGR234. Primers TSnoeE1 and TSnoeE3 were used to detect and amplify the *noeE* gene in representative 'cowpea group' isolates using the same conditions described above. These primers were also applied to template DNA originating from the recent study of Stępkowski et al. (2007) to determine the extent to which *noeE* is distributed among the different *nodA* clades of *Bradyrhizobium* (see below). For *nodA* Clade III, DNA extracts for the following strains were used: C8, CCBAU10071, CH2310, CH2318, CH2437, CH2490, CH2498, CH2506, CH2510, CPAC-M9, F3b, FTA6, KFR22, Lcamp1, Lcamp6, Lcamp8, ORSAT6, ORSAT8, ORS88, Ppau3-41, QA4, and USDA3259. DNA extracts for strains BLUH1, BLUT1, C4, CH2391, CH2352, CH2355, CH2443, CH2438, cytisus11, FTA7, ICEA, ICEB, ICED, PL41, MCLA07, Os9, RLA08, and Zaluz80 were used to represent *nodA* Clade II. For *nodA* Clade V, DNA from strain CH2493 was used, and for Clade VII, DNA from strains CH2509, Da3-1 and Rp2-1 were used, while those for strains USDA3002, USDA3475 and USDA3001 were used to represent *nodA* Clade I. All nodulation gene PCR products were purified with the QIAquick gel extraction kit (QIAGEN) and sequenced using the original PCR primers and the ABI PRISM BigDye Terminator v3.0 Cycle Sequencing Kit on the ABI3100 Automated Capillary DNA Sequencer (Applied Biosystems).

2.5. Phylogenetic analyses

Prior to phylogenetic analyses, each of the raw sequence files were inspected and corrected, where necessary, using Chromas Lite 2.0 (Technelysium) and BioEdit version 5.0.9 (Hall, 1999). These sequences were also compared to those in GenBank (National Centre for Biotechnology Information; http://www.ncbi.nlm.nih.gov/) using *blastn* (Altschul et al., 1990) to verify gene identity. All *rrs* sequences were also analyzed with CHIMERA_CHECK version 2.7 hosted by the Ribosomal Database Project II (RDP II; http://rdp8.cme.msu.edu/html/). Multiple nucleotide alignments were compiled for our new sequences, as well as published sequences obtained from GenBank. Alignments were generated with ClustalX version 1.83 (Thompson et al., 1997) and BioEdit. For the protein-coding regions, nucleotide alignments were based on aligned amino acid sequences. For this purpose, amino acid sequences were inferred in BioEdit prior to alignment using ClustalX, and nucleotide alignments were generated from the subsequent amino acid alignments by converting amino acid sequences to nucleotides in BioEdit. To determine whether the *rrs*, *recA*, and *glnII Bradyrhizobium* gene sequences represent homogenous partitions that are combinable into a single dataset, the incongruence length difference (ILD) test (Farris et al., 1995) implemented in PAUP* version 10b (Swofford, 2002) was used on parsimony informative sites only (Darlu and Lecointre, 2002; Dolphin et al., 2000; Lee, 2001). For this purpose 1000 repartitions of the datasets were subjected to heuristic searches using 100 rounds of random sequence additions and tree bisection reconnection branch swapping.

Phylogenies were constructed based on maximum likelihood (ML) using PhyML version 2.4.3 (Guindon and Gascuel, 2003), and Neighbor-Joining (NJ; Saitou and Nei, 1987) genetic distances using PAUP*. As indicated by Modeltest (Nylander, 2004; Posada and Crandall, 1998), ML and NJ analyses of the *nodA* and *nodZ* data utilized the transversional model (TVM; Posada, 2003) with unequal base frequencies, a specific proportion of invariable sites (I) and gamma correction for among-site variation (G). ML and NJ analyses of the *noeE* data utilized Tamura and Nei's (1993) model with unequal base frequencies (TN) plus G, while the *noeI* data used Zharkikh's (1994) symmetrical model (SYM) with equal base frequencies plus G. ML and NJ analyses of the *rrs* data used the TN model, the *glnII* data used the general time reversible (GTR; Tavaré, 1986) model and the *recA* data used the transitional model (TIM; Posada, 2003), all three with I+G. ML and NJ analyses of the *rrs* + *recA* + *glnII* dataset utilized the GTR + I + G model as indicated by Modeltest. Robustness of ML and NJ phylogenies was evaluated with non-parametric bootstrap analyses (Felsenstein, 1985) using 1000 pseudoreplicates and the same best-fit model parameters.

For Bayesian inference (BI) we used MrBayes version 3b4 (Huelsenbeck and Ronquist, 2001). The Metropolis-coupled Markov chain Monte Carlo search algorithm was used with 2,000,000 generations running one cold and three heated chains and sampling every hundredth tree. Bayesian posterior probabilities were calculated using the trees sampled after burin, which was determined using visual inspection of the log-likelihood scores of the sampled trees for stationarity (see Supplementary data). All BI analyses were performed at least 3 times. As the models that best fit the *nodA* data and the individual core gene regions were all submodels of the GTR model, BI analyses of the *rrs* + *recA* + *glnII* dataset and the *nodA* data both utilized the GTR + I + G model. In addition, unlinked model parameters were used across the three gene partitions for BI analysis of the *rrs* + *recA* + *glnII* dataset. All BI analyses and ML or NJ analyses involving correction for among-site variation used eight-category gamma models. For BI, NJ, and ML analyses of the *nodA* dataset, third codon positions were excluded as they were previously shown to display significant levels of substitution saturation (Moulin et al., 2004). To determine whether the *nodA* data and the three individual core gene datasets support the same phylogenetic groupings as those inferred from the *rrs* + *recA* + *glnII* dataset, and whether the *nodZ*, *noeE*, and *noeI* datasets support the *nodA* and *rrs* + *recA* + *glnII* clades, the Shimodaira–Hasegawa (SH) test (Shimodaira and Hasegawa, 1999) as implemented in CONSEL (Shimodaira and Hasegawa, 2001) was used.

3. Results

3.1. ITS PCR-RFLP analysis

To select 'cowpea group' *Bradyrhizobium* isolates with a diverse range of taxonomic relatedness, all isolates from the six localities were compared by ITS PCR-RFLP. Of the two restriction enzymes tested, the HaeIII fingerprints revealed only two groups at the 60% similarity level (data not shown). In contrast, thirteen clusters with >70% similarity were evident among the AluI fingerprints (Fig. 1) and the overall groupings appeared to reflect geographic origin rather than legume host. Isolates in clusters 1 and 2 were predominantly from Francistown and Roodeplaat, while those in 3, 4, and 5 were predominantly from Rasesa. Clusters 1–5 formed a single group (ITS-A) at the 68% similarity level that included inoculant strain CB756 from Zimbabwe. At the 77% similarity level, another group (ITS-B) contained clusters 6 (mostly from Gaborone), 7 (all from Maun) and 8 (two Roodeplaat isolates). A third group (ITS-C, 72% similarity level) contained *B. japonicum* type strain LMG6138, as well as cluster 9 (consisting of *B. elkanii* type strain LMG6134 and Rasesa isolate R5) and clusters 10–12 that contained Good Hope isolates. Cluster 12 also contained Gaborone isolates. The last group, ITS-D, included cluster 13 representing isolates from Rasesa that all shared less than 50% similarity with the rest of the isolates. Isolates from each of the 13 clusters were selected for further analysis (Table 1).

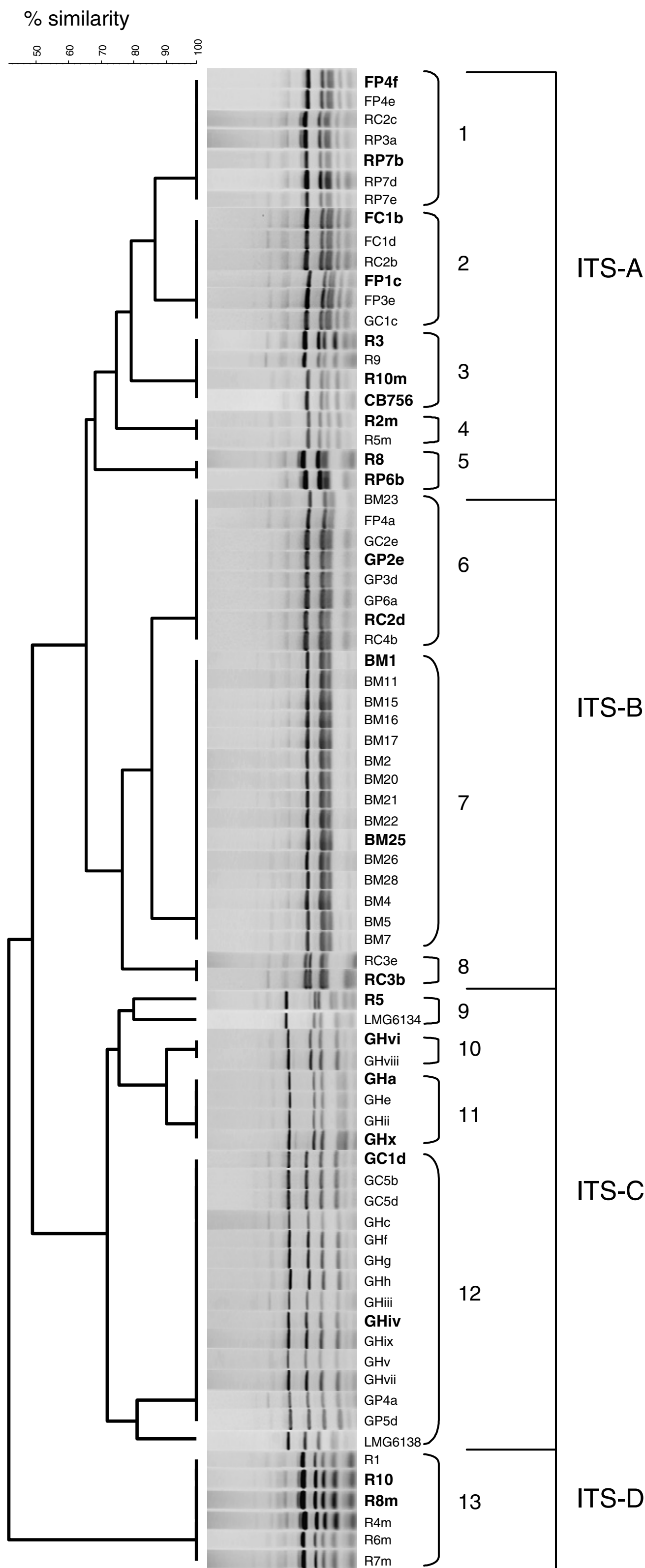


Fig. 1. UPGMA-dendrogram derived from PCR-RFLP fingerprints of *Bradyrhizobium* reference strains and isolates obtained from cowpea and peanut nodules in Botswana and South Africa generated by AluI digestion of amplified rRNA ITS PCR products. The thirteen clusters (1–13) and four major groups (ITS-A to ITS-D) are indicated to the left and the strains selected for sequencing are shown in bold. Scale shows percentage similarity.

Overall, the ITS-based fingerprint clusters did not closely match the relationships among the isolates based on the *rrs* + *recA* + *glnII* data (see below). Although, lineages SA-3 and SA-2 were restricted to groups ITS-A and ITS-B and to group ITS-C, respectively, the *B. yuanmingense* isolates were located in both the ITS-A and ITS-D groups. In some cases, isolates representing different *rrs* + *recA* + *glnII*-based lineages displayed similar ITS-RFLP profiles, suggesting that this method for detecting groups of like isolates underestimates the true diversity among the isolates studied.

3.2. Phylogenetic analyses using rrs, recA, and glnII gene sequences

Comparison of the *rrs* sequences generated in this study to those in GenBank, confirmed that the 'cowpea group' of isolates form part of the genus *Bradyrhizobium*. In an attempt to determine their relatedness to the known species and lineages of the genus, we generated an alignment including *rrs* sequences for our 'cowpea group' isolates, the *Bradyrhizobium* type strains, as well as representatives for most of the other genera of the *Bradyrhizobiaceae* family (Garrity et al., 2005b). The alignment also included representatives for the *rrs* PCR-RFLP-based genospecies reported by Lafay and Burdon (1998, 2001), the *rrs* sequence-based genospecies identified by Weir et al. (2004) and Lafay and Burdon (2006), and those identified by Vinuesa et al. (2005b) using protein-coding gene sequences. The *rrs* dataset consisted of 1357 aligned sites, seven of which were alignment gaps. However, a well-resolved phylogeny was not generated (results not shown), which was to be expected as this region is known to lack sufficient phylogenetic information for resolving the relationships among most *Bradyrhizobium* species (e.g. van Berkum and Fuhrmann, 2000; Vinuesa et al., 2005a; Willems et al., 2001b). These data did, however, show that none of the 'cowpea group' isolates appeared to be related to *B. betae* (Rivas et al., 2004) and *B. denitrificans* (van Berkum et al., 2006) and as representative *recA* and *glnII* sequences are not available in GenBank, they were not included in the other phylogenetic analyses conducted in this study. The high degree of *rrs* sequence similarity among species also complicated the detection of possible chimeric sequences that potentially resulted from recombination events between different *Bradyrhizobium* lineages as has been seen for some other rhizobial rRNA genes (Parker, 2001; van Berkum et al., 2003). Analysis with CHIMERA_CHECK did, however, show that all parts of each *rrs* sequence share significant similarity with *Bradyrhizobium* sequences only.

The *recA* and *glnII* alignments consisted of 590 and 601 aligned nucleotide positions, respectively. In addition to the isolates examined in this study, these alignments also included all bradyrhizobial taxa for which sequence information for *rrs*, *recA*, and *glnII* are available in GenBank. ML analyses of all three datasets generated trees with some similar groupings (details are shown as Supplementary data), which supported the clades representing *B. elkanii* and SA-3 identified in this study (see below, Fig. 2). The *recA* and *glnII* datasets statistically supported the *B. yuanmingense* and *B. canariense* clades, although the latter was also recovered from but not supported by the *rrs* data. The *B. japonicum* clade was supported by the *recA* data only. A clade containing members of lineage SA-1 identified in this study (see below, Fig. 2) together with *Bradyrhizobium* genospecies α (strain BC-C1), was recovered from the *recA* and *rrs* data, but was statistically supported by the *rrs* data only. The relative positions of most of these clades and lineages varied considerably between the *rrs*, *recA*, and *glnII* single-gene trees. Overall, the "backbone" of each tree generally also lacked significant bootstrap support.

The results of all three pair wise ILD tests indicated that none of the three loci represent homogenous data partitions as *P*-values of 0.01 (*recA* + *glnII*), 0.02 (*rrs* + *recA*), and 0.04 (*rrs* + *glnII*) were obtained. However, the ILD test has been shown to be associated with an excessively high type I statistical error rate (Barker and Lutzoni, 2002; Cunningham, 1997) and it has been suggested that *P*-values

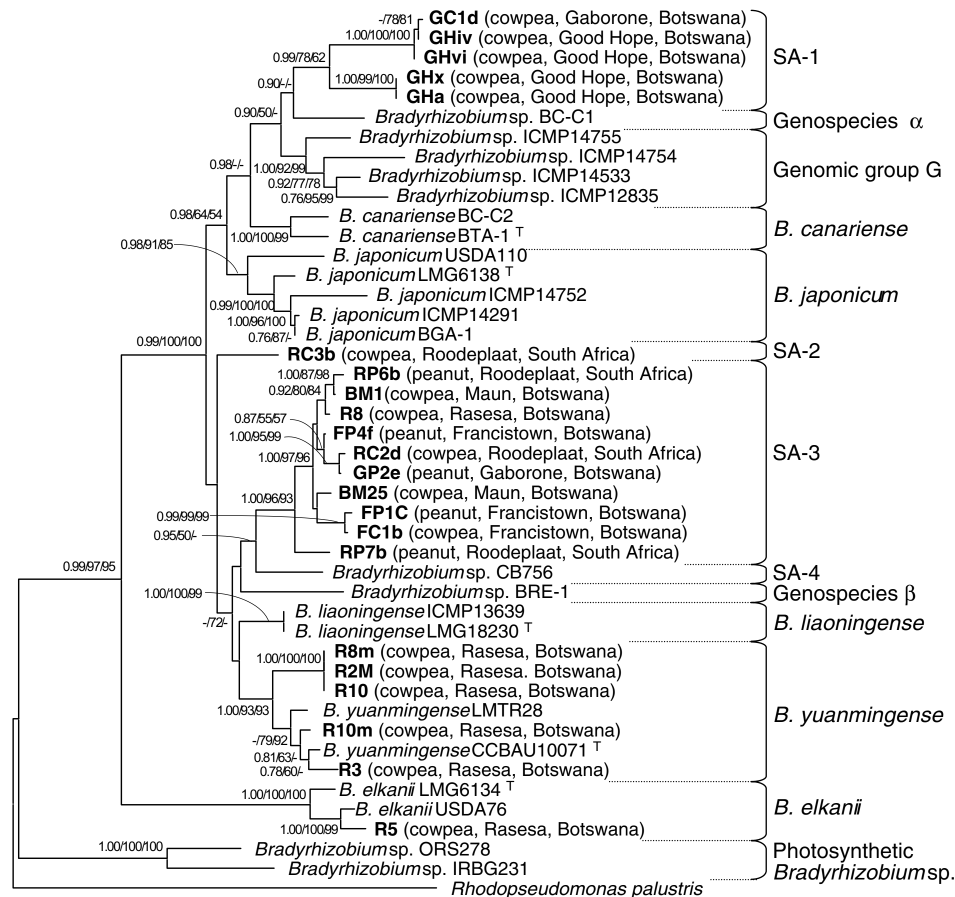


Fig. 2. A maximum likelihood (ML) phylogeny for the genus *Bradyrhizobium* based on combined *rrs*, *recA*, and *glnII* nucleotide sequence data. The southern African 'cowpea group' isolates are indicated in bold and all of the major *Bradyrhizobium* lineages are indicated to the right. *Bradyrhizobium* type strains are indicated with T. Genospecies α and β refer to those reported by Vinuesa et al. (2005b), while genomic group G was reported by Weir et al. (2004). Similar trees were generated using Bayesian inference (BI) and Neighbor-Joining (NJ) distance methods. Bootstrap values for the ML and NJ analyses, as well as Bayesian posterior probabilities are indicated at the internodes in the order BI/ML/NJ. The tree is rooted with *Rhodopseudomonas palustris*.

between 0.01 and 0.001 (making our *P*-values borderline) should be taken as a more appropriate criterion for rejection of combinability rather than the generally accepted 0.05 level. Previous research (Barker and Lutzoni, 2002; Darlu and Lecointre, 2002) has also shown that tests of combinability of data are strongly influenced by the presence of lineage-specific and site-specific heterogeneity. Accordingly, removal of the taxa with variable positions in our single-locus trees (e.g. *Bradyrhizobium* sp. strains IRBG231, BRE-1, and CB756, and *B. japonicum* USDA122), generated *P*-values of 0.07 for the *recA* + *glnII* dataset, 0.26 for the *recA* + *rrs* dataset and 0.15 for the *glnII* + *rrs* dataset. Finally, previous studies (Nixon and Carpenter, 1996; Wenzel and Siddall, 1999) have shown that combination of heterogeneous data (like our *rrs*, *recA*, and *glnII* sequences) may increase the accuracy of a phylogeny because combination of multiple datasets with different levels of homoplasy (i.e. 'phylogenetic noise') enhances the underlying phylogenetic signal thus resulting in overall more robust phylogenies. We therefore followed a 'total evidence approach' based on conditional data combination (Bull et al., 1993; Huelsenbeck et al., 1996; Kluge, 1989) and proceeded to combine the datasets for phylogenetic analyses.

Similar trees were generated using ML, NJ, and BI analyses of our *rrs* + *recA* + *glnII* dataset (Fig. 2). This topology was also not rejected ($P > 0.05$) by any of the three single-gene datasets based on our SH-test results. Overall BI, ML, and NJ analysis of the sequence data separated the taxa included into thirteen distinct lineages (Fig. 2). These correspond to the described species and lineages of *Bradyrhizobium*, as well as an additional four lineages (SA-1, SA-2, SA-3, and SA-4) harboring the southern African 'cowpea group' isolates and inoculant strain. Lineage SA-1 includes all the Good Hope isolates as well as one of the Gaborone isolates (GC1d). Lineage SA-3 include all of the Francistown and Maun isolates, three from Roodeplaat (RC2d, RP6b, and RP7b), one from Gaborone (GP2e) and one from Rasesa (R8). Roodeplaat isolate RC3b and the inoculant strain CB756 from Zimbabwe appear to represent unique lineages (SA-2 and SA-4). The majority of the isolates from Rasesa (R2m, R3, R8m, R10, and R10m) form part of the *B. yuanmingense* clade, although isolate R5 most probably represents *B. elkanii* as it forms part of the clade containing known members of this species.

3.3. Phylogenetic analyses using nodA nodZ, noeE, and noeI gene sequences

The *noeE* primers designed in this study amplified a portion of this gene in the selected 'cowpea group' strains (i.e. BM25, FC1b, GC1d, GHa, R3, R2m, RC3b, and RP6b) and the commercial inoculant strain CB756. It also amplified the *noeE* region from the DNA extracts of the four *nodA* Clade III strains (USDA3259, F3b, KFR22, and CCBAU10071), as well as from DNA extracts of two strains (CH2509 and Rp2-1) representing Clade VII. The *noeE* primer set did not allow amplification of the corresponding region of the gene in any of the other representative samples included. This may be attributable to the specific nature of the primers, thus not allowing

amplification of more variable versions of the gene. However, these primers target sequences that are sufficiently conserved to successfully amplify the *noeE* gene even from unrelated bacteria such as *Sinorhizobium* sp. NGR234. It is therefore unlikely that lack of amplification result from our primers not being able to amplify all potential *noeE* targets. Comparison of the deduced amino acid sequences for each of the *noeE* fragments sequenced, revealed that two of the examined isolates (CH2509 and the *B. japonicum* type strain) carry mutations (i.e. insertions or deletions that cause frameshits) that would inactivate the gene. Whether the remaining isolates that do not harbor such mutations indeed produce functional NoeE remains to be determined as we sequenced only part of the gene.

ML, NJ, and BI analyses of our *nodA* dataset that includes the sequences generated in this study, as well as published *Bradyrhizobium nodA* sequences (Chaintreuil et al., 2001; Chen et al., 2003; Kalita et al., 2006; Kaneko et al., 2002; Lafay and Burdon, 2006; Moulin et al., 2004; Stępkowski et al., 2007, 2005; Weir, 2006; Weir et al., 2004), produced a well-supported phylogeny (Fig. 3). The *nodA* phylogeny consisted of seven distinct clades that were previously recognized and designated as Clades I–VII (Moulin et al., 2004; Stępkowski et al., 2007, 2005). Clade III is further subdivided into Clades III.1, III.2, and III.3 (Stępkowski et al., 2007), with all of the 'cowpea group' *Bradyrhizobium* isolates, as well as the inoculant strain CB756, residing in Clade III.3. Within this clade, the 'cowpea group' isolates were included in eight well-supported groups (A–H). The only exception was Group E consisting of the *nodA* sequences of Roodeplaat isolate RC3b and the inoculant strain CB756, which lacked significant Bayesian posterior probability and bootstrap support values (Fig. 3). The composition of *nodA* Group A corresponds to that of lineage SA-3 (Fig. 2), except that the isolates from Maun form a separate group (Group D, Fig. 3). In contrast, some of the lineages identified using the *rrs* + *recA* + *glnII* dataset are split into several unrelated *nodA* groups. For example, members of lineage SA-1 (Fig. 2) are separated into *nodA* Groups C and G (Fig. 2), while the cowpea isolates identified as *B. yuanmingense* were divided between Groups B and F (Fig. 3). Interestingly, the 'cowpea group' isolates that were obtained from the root nodules of peanut are restricted to Group A.

The *nodA*, *nodZ*, *noeE*, and *noeI* datasets supported some similar groupings, although the overall topology of the individual trees did not match (compare Figs. 3 and 4). Also, neither of the *noeE* and *noeI* datasets supported the *nodA* clades inferred in this study as these hypotheses were strongly rejected by the results of our SH-tests (Table 2). These tests could, however, not reject the hypothesis that the *nodZ* dataset support similar groupings to those inferred from the *nodA* data (Table 2). Furthermore, none of the individual nodulation genes supported phylogenetic groupings similar to those inferred from the *rrs* + *recA* + *glnII* dataset (Fig. 2 and Table 2).

4. Discussion

4.1. A multigene phylogeny for Bradyrhizobium

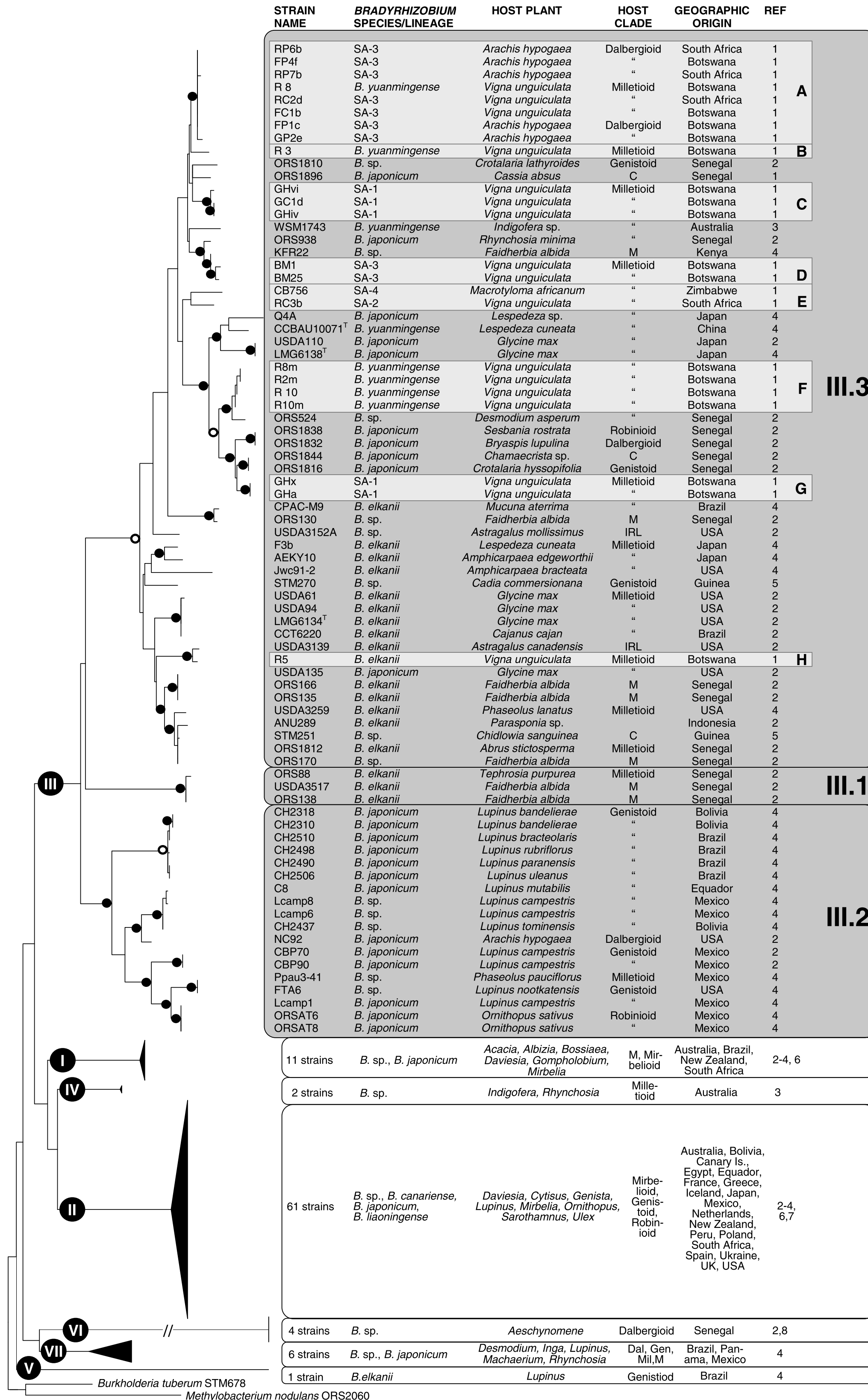
To determine the identities of the 'cowpea group' isolates of *Bradyrhizobium* nodulating cowpea and peanut from different locations in South Africa and Botswana, we used the sequence information for three loci. Single-gene phylogenies are extremely sensitive to unequal evolutionary rates among taxa and among nucleotide sites within the gene, as well as phylogenetic tree-building artefacts (Nichols, 2001). Some genes may also lack phylogenetic information for resolution of all the relationships (Gadagkar et al., 2005). Thus analyses of single-gene data for classification purposes seldom provide well-supported resolution among taxa (e.g., Gaunt et al., 2001; Nichols, 2001; Vinuesa et al., 2005a, b), while the analysis of concatenated datasets seem to produce more robust evolutionary trees (e.g. Gadagkar et al., 2005; Vinuesa et al., 2005a,b). This was also the case for our single-gene trees (see Supplementary data) and concatenated tree (Fig. 2) as the combined phylogenetic information incorporated in the *rrs*, *recA*, and *glnII* sequences outweighed that of the single loci and allowed the generation of a relatively well-supported and robust *Bradyrhizobium* phylogeny.

Our *rrs* + *recA* + *glnII* phylogeny (Fig. 2) is in agreement with the current taxonomic structure of *Bradyrhizobium* (Sawada et al., 2003; Vinuesa et al., 2005a,b; Willems et al., 2003), supporting previous findings on the increased accuracy of concatenated gene trees (Cummings and Meyer, 2005; Gadagkar et al., 2005). We show that the eight existing clades or lineages recognized in the well-resolved *Bradyrhizobium* phylogeny constructed by Vinuesa et al. (2005b) should be expanded to include the four 'cowpea group' lineages reported here, as well as the genomic group G bradyrhizobia isolated from New Zealand legumes by Weir et al. (2004). Also, our tree clearly shows that *B. japonicum*, *B. canariense*, *B. yuanmingense*, *B. liaoningense* and genospecies α and β, together with genomic group G and our novel 'cowpea group' lineages form a well-supported cohesive cluster representing a sistergroup to *B. elkanii*. This sistergroup relationship is consistent with the results of previous DNA–DNA hybridization studies showing that the genomic content of *B. elkanii* differs markedly from those of other *Bradyrhizobium* species (Willems et al., 2001b). Consistent with these studies, our data also place the photosynthetic strains at the base of the *Bradyrhizobium* phylogeny and supports the idea that members of *Bradyrhizobium* are descended from photosynthetic ancestors, possibly related to *Rhodopseudomonas* (Jarvis et al., 1986). However, further clarification of the latter issue and resolution of the exact relationships among all lineages in the genus is required because the *rrs* + *recA* + *glnII* dataset, like those previously reported (Vinuesa et al., 2005a,b), does not appear to contain adequate information for resolving the relationships among all groups in the genus. Our taxon sampling was also relatively limited (only taxa for which all three sequences were available were included) and could have contributed to the lack of resolution (Cummings and Meyer, 2005). Unequivocal reconstruction of the evolutionary history of the genus *Bradyrhizobium* will undoubtedly involve the inclusion of all possible lineages of this taxon and utilization of sequence information for additional core genes to increase the phylogenetic information content of the datasets from which phylogenies are inferred.

4.2. Diverse Bradyrhizobium species nodulate cowpea and peanut

Most of the 'cowpea group' isolates examined in this study represent novel lineages of *Bradyrhizobium* (Fig. 2). The majority of these form part of lineage SA-3, including isolates from the root nodules of cowpea grown in soils from Maun, Rasesa, and Francistown in Botswana and Roodeplaat in South Africa, as well as peanut nodules originating from Roodeplaat, Francistown, and Gaborone (Botswana) soils. Formal description of lineage SA-3 as a new species of *Bradyrhizobium*, will require additional population studies to characterize its biological properties and clearly delineate its species boundaries. The current data suggests that lineage SA-3 is found in soils ranging from loamy sand to sandy clay loam with pH between 6.2 and 6.9 (Law et al., 2007). Most of the members of this lineage appear to be capable of nodulating and establishing an effective symbiosis with both cowpea and peanut (Law et al., 2007). Lineage SA-3 also appears to be very diverse, as none of the ten SA-3 isolates included in the *rrs* + *recA* + *glnII* tree shared *recA* and *glnII* genotypes. The large variety of ITS PCR-RFLP profiles (Fig. 1) and genomic fingerprints (Law et al., 2007) generated for these isolates also suggest substantial diversity.

All of the ‘cowpea group’ isolates originating from cowpea root nodules induced in uninoculated soils from Good Hope and Gaborone in Botswana, represent lineage SA-1 (Fig. 2). This separation is supported by the ITS PCR-RFLP analysis (Group ITS-C; Fig. 1) and previous genomic fingerprinting (Law et al., 2007). Law et al. (2007) demonstrated that lineage SA-1 isolates from Good Hope appear to be adapted to cowpea as they are all incapable of inducing effective nodules on peanut. Identification and characterization



of additional isolates related to SA-1, will facilitate a better understanding of the host range and biology of this novel lineage.

Five of the seven 'cowpea group' isolates obtained from the root nodules of cowpea plants grown in soil from Rasesa in Botswana appear to represent *B. yuanmingense* (Fig. 2). The *B. yuanmingense* clade includes the type strain (CCBAU10071) for this species that was originally isolated from *Lespedeza cuneata* in Beijing, China (Yao et al., 2002), as well as *B. yuanmingense* strain LMTR28 that was isolated from *Indigofera hirsuta* in Peru (So et al., 1994; Vinuesa et al., 2005b). The *rrs* + *recA* + *glnII* data separate *B. yuanmingense* into two smaller groups with one containing CCBAU10071, LMTR28 and two Rasesa isolates (R3 and R10m), and the other, three of the Rasesa isolates (R2m, R8m and R10). As indicated by a *recA* + *glnII* tree (results not shown), the latter subgroup also includes another known *B. yuanmingense* strain TAL760 (for which *rrs* data was unavailable) isolated from *P. lunatus* in Mexico (Ormeño et al., 2006). The available information on *B. yuanmingense* is thus expanded regarding (i) its known geographic range of China, Peru, and Mexico (Gu et al., 2006; Ormeño et al., 2006; So et al., 1994; Vinuesa et al., 2005b) to include an African locality in Botswana; and (ii) its known host range of *P. lunatus*, *I. hirsuta*, *L. cuneata*, *Desmodium racemosum* and *D. fallax* (Gu et al., 2006) to include the African legume, cowpea.

None of the 'cowpea group' *Bradyrhizobium* isolates examined in this study represented *B. japonicum* , *B. canariense* or *B. liaoningense*, suggesting that they may be rare in southern African soils. The lack of *B. canariense* and *B. liaoningense* could be related to the soil types, none of which were acid or alkaline as preferred by the two species, respectively (Vinuesa et al., 2005a; Xu et al., 1995). Alternatively, these other *Bradyrhizobium* species may be unable to establish effective symbioses with peanut, cowpea and possibly also related legumes. For example, the predominant root-nodule bacteria of Genistioid legumes are mostly *B. japonicum* and *B. canariense*, while Milletioid legumes appear to be more commonly nodulated by *B. yuanmingense*, *B. elkanii* and the various lineages reported here and in previous studies (Moulin et al., 2004; Stępkowski et al., 2007, 2005; Vinuesa et al., 2005a).

Ou data conclusively show that the widely used inoculant for cowpea and peanut, CB756, is unrelated to the other 'cowpea group' isolates examined in this study. Based on the *rrs* + *recA* + *glnII* data, CB756 represents a unique and, as yet, single-taxon lineage (SA-4) within the *Bradyrhizobium* clade (Fig. 2), which may reflect its original isolation from an unrelated host, *Macrotyloma africanum*, in Zimbabwe. Previous *rrs*-based studies of elite rhizobial commercial strains identified this strain as *B. japonicum* and grouped it with the inoculant strains for peanut, *Stylosanthes* species, *Crotalaria spectabilis* and *Calopogonium* (Germano et al., 2006; Menna et al., 2006). Our results, however, clearly show that CB756 is not directly related to *B. japonicum* or its close relative *B. canariense*. Furthermore, phylogenetic analyses of our *nodA* data suggest that strain CB756 and Roodeplaat isolate RCb3 potentially shared an ancestral homolog of the gene, although this association lacks significant bootstrap support (Fig. 3).

4.3. The nodA genes of African Bradyrhizobium species share a common ancestor

Phylogenetic analyses of the *nodA* sequences of the *Bradyrhizobium* isolates from Botswana and South Africa separated the members of this genus into the seven well-supported Clades I–VII (Fig. 3) (Moulin et al., 2004; Stępkowski et al., 2007). The overall topology of the *nodA* tree corresponds to that presented by Moulin et al. (2004) who demonstrated that these sequences form a monophyletic lineage, suggesting that HGT of nodulation genes between *Bradyrhizobium* and other rhizobial lineages is a rare event. In contrast, the non-monophyletic nature of the *nodA* sequences for individual *Bradyrhizobium* lineages highlights the importance of HGT among members of the genus (compare Figs. 2 and 3). Instead of species relationships, *nodA* sequences appear to reveal geographic origin or host range (Moulin et al., 2004; Stępkowski et al., 2007, 2005), with Clade I including strains isolated from Australian soils or from native Australian *Acacia* species grown elsewhere. Clade II includes strains isolated from Genisteae and serradella hosts that are mostly of European origin. Clade IV includes strains isolated from native Australian legumes. Clade V is currently represented by a single strain isolated from *Lupinus paraguariensis* in Brazil. Members of Clades VI and VII represent, respectively, the photosynthetic *Bradyrhizobium* strains and those originating from legumes of the tribes Desmodieae, Ingeae, Dalbergieae and Phaseoleae found in Central America and Mexico. Clade III comprises strains with tropical and sub-tropical origins (see below), as well as our entire set of 'cowpea group' isolates. Among these isolates, phylogenetic relationships also reflect geographic origin (Groups A–H; Fig. 3) suggesting that the observed *nodA* distribution patterns probably reflect bacterial requirements for ability to nodulate different sets of hosts native to the respective locations in South Africa and Botswana.

Compared to the other *nodA* clades, the taxa in Clade III appear to be more diverse as they are associated with longer terminal branches (Fig. 3). The diverse nature of this clade is also reflected by the host range of its members. Some Clade III sequences originate from *Bradyrhizobium* isolates obtained from the non-legume *Parasponia* species, while others were obtained from legumes in the subfamilies Caesalpinioideae (from *Cassia*, *Chamaecrista*, and *Chidlowia*) and Mimosoideae (from *Faidherbia albida*). However, most Clade III *nodA* sequences were from bradyrhizobia of plants in the major subclades of the Papilionioideae (Schrire et al., 2005; Wojciechowski et al., 2004), i.e. the Genistoid *sensu lato*, Robinioid, Dalbergioid *sensu lato*, IRL ('inverted repeat-lacking') and Millettioid *sensu lato* clades. Although many *Bradyrhizobium* strains isolated from these plant groups harbor *nodA* sequences representing the other clades, the majority of the known *nodA* sequences associated with the milletioid legumes are located in Clade III. The sequences in this clade further seem to have a pantropical distribution with representatives originating in Africa, Australia, North America, South America, Indonesia, China, Japan, and Papua New Guinea. In this respect, all known members of Clade III.2 originate from the Americas, where they are associated predominantly with genistoid hosts, while Clades III.1 and III.3

Fig. 3. A maximum likelihood (ML) *nodA* phylogeny for the genus *Bradyrhizobium*. The *nodA* Clades I–VII are indicated on the branches leading to each of the clades, while subclades III.1–III.3 are indicated on the outer right hand side of the figure. Southern African 'cowpea group' taxa are indicated with A–H. The species/lineage diagnosis for individual strains, legume host, higher order classification of hosts according to Wojciechowski et al. (2004) (Caesalpinioideae and Mimosoideae are indicated with C and M, respectively) and geographic origin are indicated to the right of the tree. References for each of the *nodA* sequences used in this study are indicated under the heading "REF" and include the following: 1, this study; 2, Moulin et al. (2004), 3, Stępkowski et al., 2005; 4, Stępkowski et al., 2007; 5, Chen et al. (2003), 6, Weir et al. (2004) and Weir 2006; 7, Kalita et al. (2006), 8, Chaintreuil et al. (2001). Similar trees were generated using Bayesian inference (BI) and Neighbor-Joining (NJ) distance methods. Bootstrap values of >90% and >60% for the ML and NJ analyses, respectively, as well as Bayesian posterior probabilities of 1.00 are indicated at the internodes with closed black circles. Branches lacking NJ bootstrap support, but that have >90% ML bootstrap support and Bayesian posterior probabilities of >0.99 are indicated by open circles. All branches leading to each of the seven *nodA* clades received >90% and >60% for the ML and NJ bootstrap support and Bayesian posterior probabilities of 1.00. The tree is rooted with the *nodA* sequences of *Burkholderia tuberum* and *Methylobacterium nodulans*. See the Supplementary data for information on taxa that are represented in the tree by triangles.

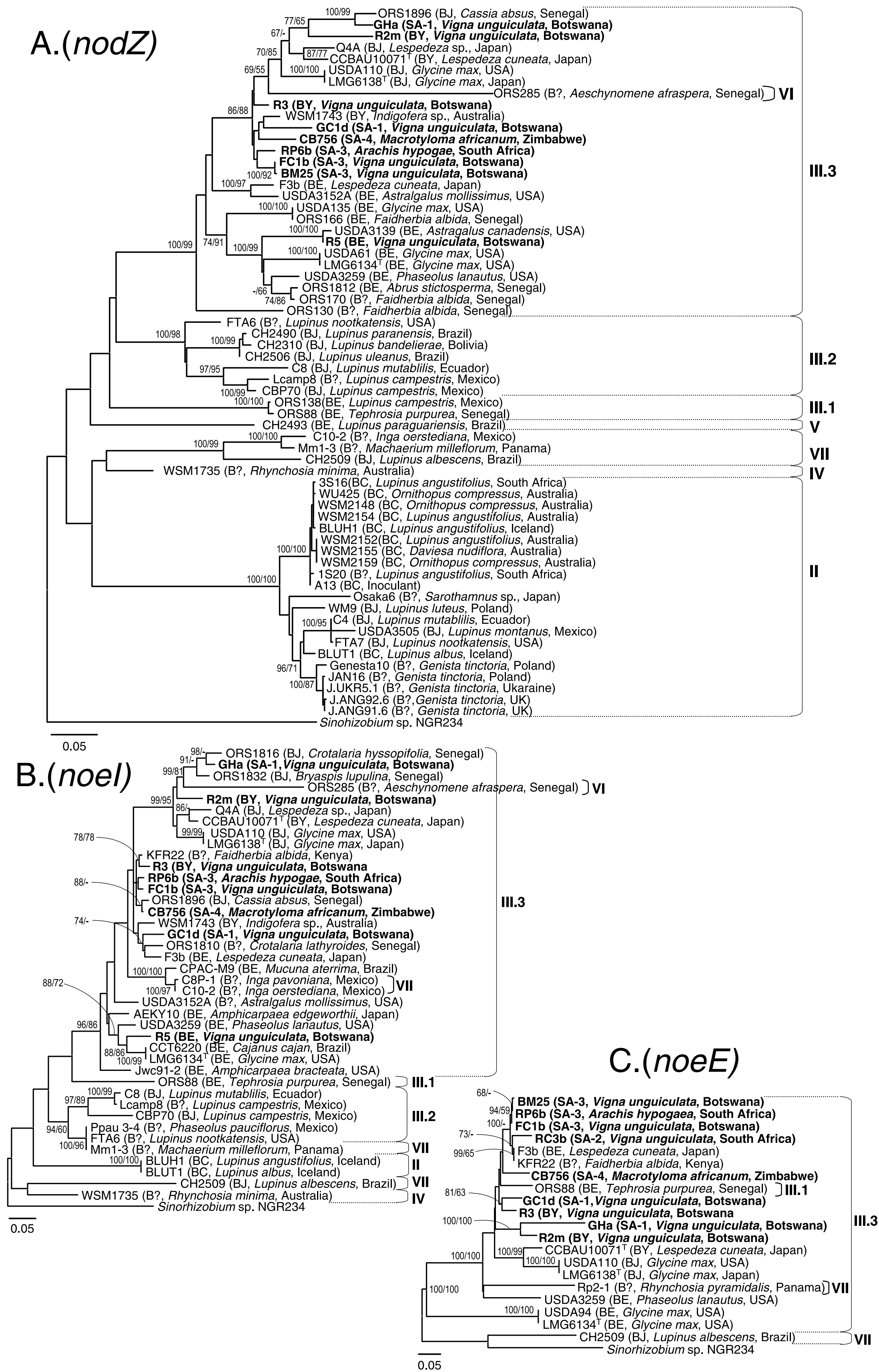


Fig. 4. Maximum likelihood (ML) phylogenies for *nodZ* (A), *noeI* (B), and *noeI* (C) gene sequences. *Bradyrhizobium* species identities, host species and geographic origin are indicated in parentheses and sequences generated for strains examined in this study are indicated in bold. *Bradyrhizobium* species identities are indicated as follows: BJ, *B. japonicum*; BC, *B. canariense*; BE, *B. elkanii*; BY, *B. yunamingense*; BL, *B. liaoningense*; B?, *Bradyrhizobium* sp.; and SA-1 to SA-3, lineages identified in this study (see Fig. 2). Major *nodA* clade and subclade groupings are indicated to the right of each tree. Similar trees were inferred using Neighbor-Joining (NJ) distance methods and bootstrap values are indicated at the branches in the order ML/NJ. The trees are rooted with corresponding sequences from *Sinorhizobium* sp. NGR234.

Table 2
Comparison of alternative *Bradyrhizobium* nodulation gene tree topologies using the SH-test[a]

Dataset tested [b]	Groups supported by the dataset tested [c]			Groups based on the *rrs* + *recA* + *glnII* phylogeny [d]			Groups based on the *nodA* phylogeny [e]		
	-lnL	△-lnL	*P*	-lnL	△-lnL	*P*	-lnL	△-lnL	*P*
nodA	12107.70	0.00	1.000	16735.72	4628.02	0.000*	–	–	–
nodZ	9664.27	0.00	1.000	12024.02	2359.75	0.000*	9735.46	71.19	0.410
noeE	5392.50	0.00	1.000	5623.83	231.33	0.000*	5899.34	506.85	0.000*
noeI	4981.32	0.00	1.000	5280.33	299.01	0.000*	5108.85	127.54	0.001*

[a] Alternative topologies were evaluated using the tests incorporated in CONSEL (Shimodaira and Hasegawa, 2001). Significantly worse ($P < 0.05$) topologies are indicated with asterisks. Similar results were obtained with the other tests included in this package, i.e. non-weighted and weighted Kishino–Hasegawa (Kishino and Hasegawa, 1989), weighted SH (Shimodaira and Hasegawa, 1999; Buckley et al., 2001), Approximately Unbiased (Shimodaira, 2002) and Bootstrap Probability (Felsenstein, 1985) tests.

[b] Datasets are the same as those used for inferring individual nodulation gene trees (Fig. 4).

[c] Tree topologies were the same as those inferred from the individual nodulation gene datasets (Fig. 4), but included only the well-supported deeper nodes with all other non-supported nodes and the terminal clades collapsed to polytomies.

[d] Test topologies were based on the *Bradyrhizobium* species groupings (Fig. 2) and defined for each dataset as follows: (Outgroup, ((*B. japonicum* isolates), (*B. elkanii* isolates), (*B. yuanmingense* isolates), (*B. canariense* isolates), (SA-1 isolates), (SA-2 isolate), (SA-3 isolates), (SA-4 isolate), *Bradyrhizobium* sp., *Bradyrhizobium* sp., etc.)).

[e] Test topologies were based on the *Bradyrhizobium nodA* Clades (Fig. 3) and defined for each dataset as follows: (Outgroup, (Clade II isolates), ((Clade III.1 isolates), (Clade III.2 isolates), (Clade III.3 isolates), (Clade IV isolates), (Clade V isolates), (Clade VI isolate), (Clade VII isolates))). "–" indicates SH-tests that were not done.

appear to predominate among *Bradyrhizobium* strains isolated in sub-Saharan Africa. Indeed, Clades III.1 and III.3 include all *nodA* sequences isolated in this and other studies from Botswana, Guinea, Kenya, Senegal, and South Africa (Chen et al., 2003; Moulin et al., 2004; Stępkowski et al., 2007, 2005). This is in contrast with other continents (except Europe where all *nodA* sequences that have been obtained to date group in Clade II), where *nodA* sequences group in various clades (Moulin et al., 2004; Stępkowski et al., 2007, 2005).

The diverse nature of *nodA* Clade III and its deep phylogenetic divergences (Fig. 3), as well as the wide geographic distribution and broad host range of its members, suggest that it might be much older than other *nodA* clades. It is therefore possible that *Bradyrhizobium* strains harboring Clade III-type *nodA* genes spread with their hosts into other parts of the world during the Eocene, following the late Cretaceous or early Tertiary origin of legumes possibly north of the Tethys Sea (Doyle and Luckow, 2003; Schrire et al., 2005; Sprent, 2007). Although such a scenario of co-dispersal would explain current *nodA* distribution patterns in Africa, North and South America, Indonesia, Japan, China, Australia, and Papua New Guinea, it does not explain the absence of *nodA* Clade III in Europe. Then during the Holocene, formation of the Sahara desert in Africa (Kuper and Kröpelin, 2006) most likely prevented further movement of Clade III *nodA* sequences to other parts of the world. As a result, all *nodA* sequences from indigenous African *Bradyrhizobium* strains and legume hosts are limited to Clade III, where their diversity is probably linked to the varied niches these bacteria inhabit and the diverse legumes they nodulate. Apart from geographical isolation, processes related to ecological speciation or drift (Rundle and Nosil, 2005) that are dependent on specific legumes growing in specific localities potentially also play important roles in the diversification of nodulation genes.

4.4. NF ornamentation and HGT of hsn genes

To examine the involvement of HGT in the evolutionary history of the nodulation genes associated with host-specificity in *Bradyrhizobium*, we also analyzed the *hsn* genes *nodZ*, *noeE*, and *noeI*. Whereas the acyl transferase NodA is responsible for transferring a fatty-acid moiety to the non-reducing sugar of the NF core, NodZ, NoeE, NoeI modify the reducing end. The activity of these and other *hsn* gene products therefore determine the overall structure of the NF core, thus conferring specificity towards different plant hosts (Carlson et al., 1993; Lopez-Lara et al., 1996; Perret et al., 2000; Ritsema et al., 1996; Roche et al., 1996; van Rhijn and Vanderleyden, 1995). Previous studies have shown that nodulation genes are most probably 'co-transferred' via transmission of the entire symbiotic island and by descent within a specific *Bradyrhizobium* lineage (Moulin et al., 2004), thus resulting in clusters of isolates with similar nodulation genes. This was true to some extent for our *nodA* and *nodZ* data as both supported clades consisting of similar isolates (Figs. 3 and 4, Table 2). However, the *noeE* and *noeI* data strongly rejected the *nodA* and *rrs* + *recA* + *glnII* phylogenies (Table 2), suggesting that *noeE* and *noeI* have their own unique and independent evolutionary origins. This seems likely as the *noeI* and *noeE* genes in the symbiotic island of *B. japonicum* USDA110 are separated by 30 and 48 kbp, respectively, from the *nodA*–*nodZ* operon (Kaneko et al., 2002). Therefore, HGT of single genes (not only the whole island) may have played an important role in shaping nodulation gene diversity and associated nodulation abilities in the ancestral *Bradyrhizobium* lineages.

All three of the examined *hsn* genes (*nodZ*, *noeE*, and *noeI*) were detected among the 'cowpea group' *Bradyrhizobium* isolates studied (Fig. 4). The gene encoding NodZ has been detected in all other previously examined strains belonging to *nodA* Clades II, III, IV, V, and VII (Moulin et al., 2004; Stępkowski et al., 2007, 2005), indicating that 6-*O*-fucosylation of the NF reducing end (Lopez-Lara et al., 1996; Quesada-Vincens et al., 1997; Quinto et al., 1997) is a common modification in this rhizobial genus. Similarly, the *noeI* gene is widespread in the *nodA* Clade III of *Bradyrhizobium* and is also found in all other examined African isolates (Moulin et al., 2004; Stępkowski et al., 2007, 2005), suggesting that the fucosyl group of African NFs are frequently 2-*O*-methylated (Jabbouri et al., 1998). The *noeE* gene encodes a fucose-specific sulfotransferase responsible for 3-*O*-sulfation of the NF (Hanin et al., 1997; Quesada-Vincens et al., 1998), and its presence among African 'cowpea group' strains is in line with previous findings that the NFs of *Bradyrhizobium* strains associated with *F. albida* nodules in Senegal are fucosylated and partially sulfated (Ferro et al., 2000). Accordingly, we also detected this gene among the previously reported *nodA* Clade III.3 strains, i.e. strains KFR22 from *F. albida* in Kenya, F3b and CBBAU10071 from *L. cuneata* in Japan, and ORS88 from *Tephrosia purpurea* in Senegal (Moulin et al., 2004; Stępkowski et al., 2007). The *noeE* gene was also present in two representatives for *nodA* Clade VII (strain CH2509 from *Lupinus albescens* in Brazil and Rp2-1 from *Rhynchosia pyramidalis* in Panama). Clarification of the exact role of the *noeE* gene in host range (if any) is required, especially the relationship between NF structure and symbiotic effectiveness on a range of native and non-native hosts in geographic regions where this gene occurs. However, our findings imply that highly ornamented NFs could be a common feature of African *Bradyrhizobium* strains, and may be an important adaptation facilitating nodulation of a broad range of the legume species native to this legume-rich continent.

5. Conclusions

We studied the evolution of 'cowpea group' isolates of *Bradyrhizobium* nodulating cowpea and peanut from different locations in South Africa and Botswana. Analyses of multiple loci allowed us to account for the independent histories of different loci and the mosaic nature of the bacterial genome (Ochman et al., 2000; Gogarten et al., 2002; Koonin et al., 2001; Welch et al., 2002). Using the combined information for genes encoded on the core genome, we showed that these bacteria are diverse, representing a number of distinct *Bradyrhizobium* groups, the majority of which require further characterization in order to be described as new species. Analyses of nodulation genes that form part of the accessory genome, revealed the presence of several groups that are apparently locally adapted and/or regionally restricted. The combined effects of vertical and horizontal gene transfer therefore appear to shape the evolution of the 'cowpea group' *Bradyrhizobium* species as they generally inherit their core genes from ancestral conspecific isolates, whereas they or their ancestors acquired their nodulation genes from unrelated sources.

Few previous studies used DNA-based methods to address the identity and diversity of the root-nodule bacteria associated with cowpea and its relatives. Krasova-Wade et al. (2003), Zhang et al. (2007) and Yokoyama et al. (2006) used *rrs* sequences to show that diverse *Bradyrhizobium* isolates are associated with different cultivars of cowpea in Senegal and China, and *Vigna* species in Thailand, respectively. Their results suggested that some isolates are closely related to known *Bradyrhizobium* species, while others potentially represent novel lineages. Yokoyama et al. (2006) also used PCR-RFLP analyses of the common *nodABC* region to show that *Bradyrhizobium* isolates obtained from a specific plant species at a specific site displayed similar profiles. The overall trends in all of these previous reports mirror our findings. However, the present and other studies (e.g., Stępkowski et al., 2007; Vinuesa et al., 2005b), clearly show that the best available tool for classifying strains of *Bradyrhizobium* is application of a multilocus sequence analysis approach (Gevers et al., 2005) using core genome genes. Furthermore, we showed that sequence analyses of nodulation genes located on the accessory symbiotic genome reveal valuable information on the role of HGT and the diversity of *Bradyrhizobium* strains associated with cowpea in South Africa and Botswana, many of which form locally restricted groups. The data presented here strongly augment the growing body of information on *Bradyrhizobium* population biology, and should contribute significantly to our understanding of the forces that shape the evolution of this important group of bacteria.

Acknowledgments

This work was partially funded by the Polish Ministry of Education and Science (Grant 2P04C 063 26) and the Regional S&T Programme of the South African Department of Science and Technology.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2008.04.032.

References

Allen, O.N., Allen, E.K., 1981. The Leguminosae: A Source Book of Characteristics, Uses and Nodulation. The University of Wisconsin press, Wisconsin, USA.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic Local Alignment Search Tool. J. Mol. Biol. 215, 403–410.

Barker, F.K., Lutzoni, F.M., 2002. The utility of the incongruence length difference test. Syst. Biol. 51, 625–637.

Barrett, R.P., 1990. Legume species as leaf vegetables. In: Janick, J., Simon, J.E. (Eds.), Advances in New Crops. Timber Press, Portland, Oregon, USA, pp. 391–396.

Botha, W.J., Bloem, J.F., Law, I.J., 2002. *Bradyrhizobium* sp. (*Lupinus*) in the winter rainfall region of South Africa. Biol. Fert. Soils 36, 335–343.

Botha, W.J., Jaftha, J.B., Bloem, J.F., Habig, J.H., Law, I.J., 2004. Effect of soil bradyrhizobia on the success of soybean inoculant strain 1809. Microbiol. Res. 159, 219–231.

Bressani, R., 1985. Nutritive value of cowpea. In: Singh, S.R., Rachie, K.O. (Eds.), Cowpea Research, Production and Utilization. Wiley, Winchester, UK, pp. 353–359.

Buckley, T.R., Simon, C., Shimodaira, H., Chambers, G.K., 2001. Evaluating hypotheses on the origin and evolution of the New Zealand alpine cicadas (Maoricicada) using multiple comparison tests of tree topology. Mol. Biol. Evol. 18, 223–234.

Bull, J.J., Huelsenbeck, J.P., Cunningham, C.W., Swofford, D.L., Waddell, P.J., 1993. Partitioning and combining data in phylogenetic analysis. Syst. Biol. 42, 384–397.

Carlson, R.W., Sanjuan, J., Bhat, U.R., Glushka, J., Spaink, H.P., Wijfjes, A.H., van Brussel, A.A., Stokkermans, T.J., Peters, N.K., Stacey, G., 1993. The structures and biological activities of the lipooligosaccharide nodulation signals produced by type I and type II strains of *Bradyrhizobium japonicum*. J. Biol. Chem. 268, 8372–18381.

Chaintreuil, C., Boivin, C., Dreyfus, B., Giraud, E., 2001. Characterization of the common nodulation genes of the photosynthetic *Bradyrhizobium* sp. ORS285 reveals the presence of a new insertion sequence upstream of *nodA*. FEMS Microbiol. Lett. 194, 83–86.

Chen, W.-M., Moulin, L., Bontemps, C., Vandamme, P., Bena, G., Boivin-Masson, C., 2003. Legume symbiotic nitrogen fixation by β-Proteobacteria is widespread in nature. J. Bacteriol. 185, 7266–7272.

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

Cummings, M.P., Meyer, A., 2005. Magic bullets and golden rules: Data sampling in molecular phylogenetics. Zoology 108, 329–336.

Dakora, F.D., Keya, S.O., 1997. Contribution of legume nitrogen fixation to sustainable agriculture in Sub-Saharan Africa. Soil Biol. Biochem. 29, 809–817.

Darlu, P., Lecointre, G., 2002. When does the incongruence length difference test fail? Mol. Biol. Evol. 19, 432–437.

Dénarié, J., Debellé, F., Promé, J.C., 1996. *Rhizobium* lipo-chitooligosaccharide nodulation factors: signalling molecules mediating recognition and morphogenesis. Annu. Rev. Biochem. 65, 503–535.

Dolphin, K., Belshaw, R., Orme, C.D.L., Quicke, D.L.J., 2000. Noise and incongruence: interpreting results of the incongruence length difference test. Mol. Phylogenet. Evol. 17, 401–406.

Doyle, J.J., Luckow, M.A., 2003. The rest of the iceberg. Legume diversity in a phylogenetic context. Plant Physiol. 131, 900–910.

Duke, J.A., 1981. Handbook of Legumes of World Economic Importance. Plenum Press, London, UK.

Farris, J.S., Källersjö, J., Kluge, A.G., Bult, C., 1995. Testing significance of congruence. Cladistics 10, 315–319.

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

Ferro, M., Lorquin, J., Ba, S., Sanon, K., Promé, J.C., Boivin, C., 2000. *Bradyrhizobium* sp. strains that nodulate the leguminous tree *Acacia albida* produce fucosylated and partially sulfated Nod factors. Appl. Environ. Microbiol. 66, 5078–5082.

Flores, M., Morales, L., Avila, A., González, V., Bustos, P.A., Garicía, D., Mora, Y., Guo, X., Collado-Vides Piñero, D., Dávila, G., Mora, J., Palacios, R., 2005. Diversification of DNA sequences in the symbiotic genome of *Rhizobium etli*. J. Bacteriol. 187, 7185–7192.

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

Gadagkar, S.R., Rosenberg, M.S., Kumar, S., 2005. Inferring species phylogenies from multiple genes: Concatenated sequence tree versus consensus gene tree. J. Exp. Zool. (Mol. Dev. Evol.) 304B, 64–74.

Garrity, G.M., Bell, J.A., Lilburn, T., 2005a. Class I. Alphaproteobacteria class. nov. In: Brenner, D.J., Krieg, N.R., Stanley, J.T., Garrity, G.M. (Eds.), Bergey's Manual of Systematic Bacteriology, second ed., vol. 2, (The Proteobacteria), part C (The Alpha-, Beta-, Delta-, and Epsilonproteobacteria) Springer, New York, USA, p. 1.

Garrity, G.M., Bell, J.A., Lilburn, T., 2005b. Family VII. *Bradyrhizobiaceae* fam. nov. In: Brenner, D.J., Krieg, N.R., Stanley, J.T., Garrity, G.M. (Eds.), Bergey's Manual of Systematic Bacteriology, second ed., vol. 2 (The Proteobacteria), part C (The Alpha-, Beta-, Delta-, and Epsilonproteobacteria), Springer, New York, USA, p. 438.

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

Germano, M.G., Menna, P., Mostasso, F.L., Hungria, M., 2006. RFLP analysis of the rRNA operon of a Brazilian collection of bradyrhizobial strains from 33 legume species. Int. J. Syst. Evol. Microbiol. 56, 217–229.

Gevers, D., Cohan, F.M., Lawrence, J.G., Spratt, B.G., Coenye, T., Feil, E.J., Stackebrandt, E., Van de Peer, Y., Vandamme, P., Thompson, F.L., Swings, J., 2005. Re-evaluating prokaryotic species. Nat. Rev. Microbiol. 3, 733–739.

Gogarten, J.P., Doolittle, W.F., Lawrence, J.G., 2002. Prokaryotic evolution in light of gene transfer. Mol. Biol. Evol. 19, 2226–2238.

Graham, P.H., Vance, C.P., 2003. Legumes: importance and constraints to greater use. Plant Physiol. 131, 872–877.

Gu, J., Wang, E.T., Chen, W.X., 2006. Genetic diversity of rhizobia associated with *Desmodium* species grown in China. Lett. Appl. Microbiol. 44, 286–292.

Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52, 696–704.

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/ NT. Nucl. Acids Symp. Ser. 41, 95–98.

Hanin, N., Jabbouri, S., Quesada-Vincens, D., 1997. Sulphation of *Rhizobium* sp. NGR234 Nod factors is dependent on *noeE*, a new host-specificity gene. Mol. Microbiol. 24, 1119–1129.

Huelsenbeck, J.P., Bull, J.J., Cunningham, C.W., 1996. Combining data in phylogenetic analysis. Trends Evol. Ecol. 11, 152–158.

Huelsenbeck, J.P., Ronquist, F., 2001. MrBayes: Bayesian inference of phylogenetic trees. Bioinformatics 17, 754–755.

Jabbouri, S., Relic, B., Hanin, M., Kamalaprija, P., Burger, U., Promé, D., Promé, J.C., Broughton, W.J., 1998. *nolO* and *noeI* (*HsnIII*) of *Rhizobium* sp. NGR234 are involved in 3-*O*-carbamoylation and 2-*O*-methylation of Nod factors. J. Biol. Chem. 273, 12047–12055.

Jarvis, B.D.W., Gillis, M., De Ley, J., 1986. Intra- and intergeneric similarities between the ribosomal ribonucleic cistrons of *Rhizobium* and *Bradyrhizobium* species and related bacteria. Int. J. Syst. Bacteriol. 36, 129–138.

Jordan, D., 1989. Numerical taxonomy of photosynthetic rhizobia nodulating *Aeschynomene* species. Int. J. Syst. Bacteriol. 44, 62–73.

Kalita, M., Stępkowski, T., Łotocka, B., Małek, W., 2006. Phylogeny of nodulation genes and symbiotic properties of *Genista tinctoria* bradyrhizobia. Arch. Microbiol. 186, 87–97.

Kaluza, K., Hahn, M., Hennecke, H., 1985. Repeated sequences similar to insertion elements clustered around the *nif* region of the *Rhizobium japonicum* genome. J. Bacteriol. 62, 535–542.

Kaneko, T., Nakamura, Y., Sato, S., Minamisawa, K., Uchiumi, T., Sasamoto, S., Watanabe, A., Idesawa, K., Iriguchi, M., Kawashima, K., Kohara, M., Matsumoto, M., Shimpo, S., Tsuruoka, H., Wada, T., Yamada, M., Tabata, S., 2002. Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110. DNA Res. 9, 189–197.

Kishino, H., Hasegawa, M., 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. J. Mol. Evol. 29, 170–179.

Kluge, A.G., 1989. A concern for evidence and a phylogenetic hypothesis of relationships among *Epicrates* (Boideae, Serpentes). Syst. Zool. 38, 7–25.

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

Krasova-Wade, T., Ndoye, I., Braconnier, S., Sarr, B., de Lajudie, P., Neyra, M., 2003. Diversity of indigenous bradyrhizobia associated with three cowpea cultivars (*Vigna unguiculata* (L.) Walp.) grown under limited and favourable water conditions in Senegal (West Africa). Afr. J. Biotech. 2, 13–22.

Kuper, R., Kröpelin, S., 2006. Climate-controlled Holocene occupation in the Sahara: motor of Africa's evolution. Science 313, 803–807.

Kuykendall, L.D., 2005a. Genus I. *Bradyrhizobium* Jordan 1982. In: Brenner, D.J., Krieg, N.R., Stanley, J.T., Garrity, G.M. (Eds.), Bergey's Manual of Systematic Bacteriology, second ed., vol. 2 (The Proteobacteria), part C (The Alpha-, Beta-, Delta-, and Epsilonproteobacteria), Springer, New York, USA, pp. 438–443.

Kuykendall, L.D., 2005b. Order VI. Rhizobiales ord. nov. In: Brenner, D.J., Krieg, N.R., Stanley, J.T., Garrity, G.M. (Eds.), Bergey's Manual of Systematic Bacteriology, second ed., vol. 2 (The Proteobacteria), part C (The Alpha-, Beta-, Delta-, and Epsilonproteobacteria), Springer, New York, USA, p. 324.

Lafay, B., Bullier, E., Burdon, J.J., 2006. Bradyrhizobia isolated from root nodules of *Parasponia* (Ulmaceae) do not constitute a separate coherent lineage. Int. J. Syst. Evol. Microbiol. 56, 1013–1018.

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.

Lafay, B., Burdon, J.J., 2006. Molecular diversity of rhizobia nodulating the invasive legume *Cytisus scoparius* in Australia. J. Appl. Microbiol. 100, 1228–1238.

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.

Law, I.J., Botha, W.J., Majaule, U., Phalane, F., 2007. Symbiotic and genomic diversity of cowpea bradyrhizobia from soils in Botswana and South Africa. Biol. Fert. Soils 43, 653–663.

Lee, M.S.Y., 2001. Uninformative characters and apparent conflict between molecules and morphology. Mol. Biol. Evol. 18, 676–680.

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

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

Molouba, F., Lorquin, J., Willems, A., Hoste, B., Giraud, E., Dreyfus, B., Gillis, M., de Lajudie, P., Masson-Boivin, C., 1999. Photosynthetic Bradyrhizobia from *Aeschynomene* spp. are specific to stem-nodulated species and form a separate 16S ribosomal DNA restriction fragment length polymorphism group. Appl. Environ. Microbiol. 65, 3084–3094.

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

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

Mutch, L.A., Young, J.P.W., 2004. Diversity and specificity of *Rhizobium leguminosarum* biovar *viciae* on wild and cultivated legumes. Mol. Ecol. 13, 2435–2444.

Nichols, R., 2001. Gene trees and species trees are not the same. Trends Ecol. Evol. 16, 358–364.

Nixon, K.C., Carpenter, J.M., 1996. On simultaneous analysis. Cladistics 12, 221–241.

Nylander, J.A.A., 2004. MrModeltest v2. Program distributed by the author, Evolutionary Biology Centre, Uppsala University.

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

Ormeño, E., Vinuesa, P., Zúñiga-Dávila, D., Martínez-Romero, E., 2006. Molecular diversity of native bradyrhizobia isolated from Lima bean (*Phaseolus lunatus* L.) in Peru. Syst. Appl. Microbiol. 29, 253–262.

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., Lafay, B., Burdon, J.J., van Berkum, P., 2002. Conflicting phylogeographic patterns in rRNA and *nifD* indicate regionally restricted gene transfer in *Bradyrhizobium*. Microbiology 148, 2557–2565.

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

Philippe, H., Douady, C.J., 2003. Horizontal gene transfer and phylogenetics. Curr. Opin. Microbiol. 6, 498–505.

Posada, D., 2003. Using Modeltest and PAUP* to select a model of nucleotide substitution. In: Baxevanis, A.D., Davison, D.B., Page, R.D.M., Petsko, G.A., Stein, L.D., Stormo, G.D. (Eds.), Current Protocols in Bioinformatics. John Wiley & Sons, Inc., pp. 6.5.1–6.5.14.

Posada, D., Crandall, K.A., 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14, 817–818.

Qian, J., Parker, M.A., 2002. Contrasting *nifD* and ribosomal gene relationships among *Mesorhizobium* from *Lotus oroboides* in northern Mexico. Syst. Appl. Microbiol. 25, 68–73.

Quesada-Vincens, D., Fellay, R., Nasim, T., Viprey, V., Burger, U., Promé, J.C., Broughton, W.J., Jabbouri, S., 1997. *Rhizobium* sp. strain NGR234 NodZ protein is a fucosyltransferase. J. Bacteriol. 179, 5087–5093.

Quesada-Vincens, D., Hanin, M., Broughton, W.J., Jabbouri, S., 1998. *In vitro* sulfotransferase activity of NoeE, a nodulation protein of *Rhizobium* sp. NGR234. Mol. Plant Microbe Interact. 11, 592–600.

Quinto, C., Wijfjes, A.H., Bloemberg, G.V., Blok-Tip, L., López-Lara, I.M., Lugtenberg, B.J., Thomas-Oates, J.E., Spaink, H.P., 1997. Bacterial nodulation protein NodZ is a chitin oligosaccharide fucosyltransferase which can also recognize related substrates of animal origin. Proc. Nat. Acad. Sci. USA 94, 4336–4341.

Ritsema, T., Wijfjes, A.H.M., Lugtenberg, B.J.J., Spaink, H.P., 1996. *Rhizobium* nodulation protein NodA is a host-specific determinant of the transfer of fatty acids in Nod factor biosynthesis. Mol. Gen. Genet. 251, 44–51.

Rivas, R., Willems, A., Palomo, J.L., García-Benavides, P., Mateos, P.F., Martínez-Molina, E., Gillis, M., Velázquez, E., 2004. *Bradyrhizobium betae* sp. nov., isolated from roots of *Beta vulgaris* affected by tumour-like deformations. Int. J. Syst. Evol. Microbiol. 54, 1271–1275.

Roche, P., Maillet, F., Plazanet, C., Debellé, F., Ferro, M., Truchet, G., Promé, J.-C., Dénarié, J., 1996. The common *nodABC* genes of *Rhizobium meliloti* are host-range determinants. Proc. Nat. Acad. Sci. USA 93, 15305–15310.

Rundle, H.D., Nosil, P., 2005. Ecological speciation. Ecol. Lett. 8, 336–352.

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

Sawada, H., Kuykendall, L.D., Young, J.M., 2003. Changing concepts in the systematics of bacterial nitrogen-fixing legume symbionts. J. Gen. Appl. Microbiol. 49, 155–179.

Schrire, B.D., Lewis, G.P., Lavin, M., 2005. Biogeography of the Leguminosae. In: Lewis, G., Schrire, B., Mackinder, B., Lock, M. (Eds.), Legumes of the World. Royal Botanic Gardens, Kew, London, UK, pp. 21–54.

Schultze, M., Kondorosi, A., 1998. Regulation of symbiotic root nodule development. Annu. Rev. Genet. 32, 33–57.

Shimodaira, H., 2002. An approximately unbiased test of phylogenetic tree selection. Syst. Biol. 51, 492–508.

Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16, 1114–1116.

Shimodaira, H., Hasegawa, M., 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics 17, 1246–1247.

Singh, B.B., Ajeigbe, H.A., Tarawali, S.A., Fernandez-Rivera, S., Abubakar, M., 2003. Improving the production and utilization of cowpea as food and fodder. Field Crops Res. 84, 169–177.

Singleton, P.W., Bohlool, B.B., Nakao, P.L., 1991. Legume response to rhizobial inoculation in the tropics: myths and realities. In: Lal, R., Sanchez, P. (Eds.), Myths in Science of Soils in the Tropics. ASA special publication ASA-SSSA-CSSA, Madison, Wisconsin, USA, pp. 135–155.

So, R.B., Ladha, J.K., Young, J.P., 1994. Photosynthetic symbionts of *Aeschynomene* spp. form a cluster with bradyrhizobia on the basis of fatty acid and rRNA analyses. Int. J. Syst. Bacteriol. 44, 392–403.

Somasegaran, P., Hoben, H.J., 1994. Handbook of Rhizobia. Methods in legume–rhizobium technology. Springer-Verlag.
Sprent, J.I., 2007. Evolving ideas of legume evolution and diversity: a taxonomic perspective on the occurrence of nodulation. New Phytol. 174, 111–125.
Steele, W.M., Allen, D.J., Summerfield, R.J., 1985. Cowpea (*Vigna unguiculata* (L.) Walp.). In: Summerfield, R.J., Roberts, E.H. (Eds.), Grain Legume Crops. Collins, London, UK, pp. 520–583.
Steenkamp, E.T., Wright, J., Baldauf, S.L., 2006. The protistan origins of animals and fungi. Mol. Biol. Evol. 23, 93–106.
Stępkowski, T., Hughes, C.E., Law, I.J., Markiewicz, L., Gurda, D., Chlebicka, A., Moulin, L., 2007. Diversification of lupine *Bradyrhizobium* strains: evidence from nodulation gene trees. Appl. Environ. Microbiol. 73, 3254–3264.
Stępkowski, T., Legocki, A., 2001. Reduction of bacterial genome size and expansion resulting from obligate intracellular life style and adaptation to soil habitat. Acta Biochim. Pol. 48, 367–381.
Stępkowski, T., Moulin, L., Krzyzańska, A., McInnes, A., Law, I.J., Howieson, J., 2005. European origin of *Bradyrhizobium* populations infecting lupins and serradella in soils of western Australia and South Africa. Appl. Environ. Microbiol. 71, 7041–7052.
Strijdom, B.W., 1998. South African studies on biological nitrogen-fixing systems and the exploitation of the nodule bacterium–legume symbiosis. S. Afr. J. Sci. 94, 11–23.
Sullivan, J.T., Ronson, C.W., 1998. Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into the phe-tRNA gene. Proc. Nat. Acad. Sci. USA 95, 5145–5149.
Suominen, L., Roos, C., Lortet, C., Paulin, L., Lindström, K., 2001. Identification and structure of the *Rhizobium galegae* common nodulation genes: evidence for horizontal gene transfer. Mol. Biol. Evol. 18, 907–916.
Swofford, D.L., 2002. PAUP∗: phylogenetic analysis using parsimony (software). Sinauer Associates, Sunderland, Massachusetts, USA.
Tamura, K., Nei, M., 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. 10, 512–526.
Tavaré, S., 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. In: Miura, R.M. (Ed.), Some Mathematical Questions in Biology—DNA Sequence Analysis. American Mathematical Society, Providence, Rhode Island, USA, pp. 57–86.
Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucl. Acids Res. 25, 4876–4882.
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., Terefework, Z., Paulin, L., Suomalainin, S., Lindstrom, K., Eardly, B.D., 2003. Discordant phylogenies with the *rrn* loci of rhizobia. J. Bacteriol. 185, 2988–2998.
van Berkum, P., Leibold, J.M., Eardly, B.D., 2006. Proposal for combining *Bradyrhizobium* spp. (*Aeschynomene indica*) with *Blastobacter denitrificans* and to transfer *Blastobacter denitrificans* (Hirsh and Muller, 1985) to the genus *Bradyrhizobium* as *Bradyrhizobium denitrificans* (comb. nov.). Syst. Appl. Microbiol. 29, 207–215.
van Rhijn, R., Vanderleyden, J., 1995. The Rhizobium–legume symbiosis. Microbiol. Rev. 59, 124–142.
Vinuesa, P., Leon-Barrios, M., Silva, C., Willems, A., Jarabo-Lorenzo, A., Perez-Galdona, R., Werner, D., Martínez-Romero, E., 2005a. *Bradyrhizobium canariense* sp. nov., an acid-tolerant endosymbiont that nodulates genistoid legumes (Papilionoideae: Genisteae) from the Canary Islands, along with *Bradyrhizobium japonicum* bv. *Genistearum*, *Bradyrhizobium* genospecies α and *Bradyrhizobium* genospecies β. Int. J. Syst. Evol. Microbiol. 55, 569–575.
Vinuesa, P., Silva, C., Werner, D., Martínez-Romero, E., 2005b. Population genetics and phylogenetic inference in bacterial molecular systematics: the roles of migration and recombination in *Bradyrhizobium* species cohesion and delineation. Mol. Phylogenet. Evol. 35, 29–54.
Weir, B.S., 2006. Ph.D. thesis. Systematics, specificity, and ecology of New Zealand rhizobia. School of Biological Sciences. The University of Auckland, New Zealand.
Weir, B.S., Turner, S.J., Silvester, W.B., Park, D.C., Young, J.M., 2004. Unexpectedly diverse *Mesorhizobium* strains and *Rhizobium leguminosarum* nodulate native legume genera of New Zealand, while introduced legume weeds are nodulated by *Bradyrhizobium* species. Appl. Environ. Microbiol. 70, 5980–5987.
Welch, R.A., Burland, V., Plunkett, G., III, Redford, P., Roesch, P., Rasko, D., Buckles, E.L., Liou, S.R., Boutin, A., Hackett, J., Stroud, D., Mayhew, G.F., Rose, D.J., Zhou, S., Schwartz, D.C., Perna, N.T., Mobley, H.L.T., Donnenberg, M.S., Blattner, F.R., 2002. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. Proc. Nat. Acad. Sci. USA 99, 17020–17024.
Wenzel, J.W., Siddall, M.E., 1999. Noise. Cladistics 15, 51–64.
Willems, A., Coopman, R., Gillis, M., 2001a. Comparison of sequence analysis of 16S-23S rDNA spacer regions, AFLP analysis and DNA–DNA hybridizations in *Bradyrhizobium*. Int. J. Syst. Evol. Microbiol. 51, 623–632.
Willems, A., Doignon-Bourcier, F., Goris, J., Coopman, R., de Lajudie, P., De Vos, P., Gillis, M., 2001b. DNA–DNA hybridization study of *Bradyrhizobium* strains. Int. J. Syst. Evol. Microbiol. 51, 1315–1322.
Willems, A., Munive, A., de Lajudie, P., Gillis, M., 2003. In most *Bradyrhizobium* groups sequence comparison of 16S-23S rDNA internal transcribed spacer regions corroborates DNA–DNA hybridizations. Syst. Appl. Microbiol. 26, 203–210.
Wojciechowski, M.F., Lavin, M., Sanderson, M.J., 2004. A phylogeny of legumes (Leguminosea) based on analysis of the plastid *matK* gene resolves many of the well-supported subclades within the family. Am. J. Bot. 91, 1846–1862.
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.
Yao, Z.Y., Kan, F.L., Wang, E.T., Wei, G.H., Chen, W.X., 2002. Characterization of rhizobia that nodulate legume species of the genus *Lespedeza* and description of *Bradyrhizobium yuanmingense* sp. nov. Int. J. Syst. Evol. Microbiol. 52, 2219–2230.
Yasuta, T., Okazaki, S., Mitsui, H., Yuhashi, K., Ezura, H., Minamisawa, K., 2001. DNA sequence and mutational analysis of rhizobitoxine biosynthesis genes in *Bradyrhizobium elkanii*. Appl. Environ. Microbiol. 67, 4999–5009.
Yokoyama, T., Tomooka, N., Okabayashi, M., Kaga, A., Boonkerd, N., Vaughan, D.A., 2006. Variation in the *nod* gene RFLPs, nucleotide sequences of 16S rRNA genes, Nod factors, and nodulation abilities of *Bradyrhizobium* strains isolated from Thai *Vigna* plants. Can. J. Microbiol. 52, 31–46.
Young, J.P., Crossman, L., Johnston, A., Thomson, N., Ghazoui, Z., Hull, K., Wexler, M., Curson, A., Todd, J., Poole, P., Mauchline, T., East, A., Quail, M., Churcher, C., Arrowsmith, C., Cherevach, I., Chillingworth, T., Clarke, K., Cronin, A., Davis, P., Fraser, A., Hance, Z., Hauser, H., Jagels, K., Moule, S., Mungall, K., Norbertczak, H., Rabbinowitsch, E., Sanders, M., Simmonds, M., Whitehead, S., Parkhill, J., 2006. The genome of *Rhizobium leguminosarum* has recognizable core and accessory components. Genome Biol. 7, R34.
Zhang, W.T., Yang, J.E., Yuan, T.Y., Zhou, J.C., 2007. Genetic diversity and phylogeny of indigenous rhizobia from cowpea [*Vigna unguiculata* (L.) Walp.]. Biol. Fert. Soils 44, 201–210.
Zharkikh, A., 1994. Estimation of evolutionary distances between nucleotide sequences. J. Mol. Evol. 39, 315–329.