CHAPTER 6

Phylogenetic relationship of a selection of indigenous South African rhizobia based on partial *nifH* gene sequencing and RFLP of the *nodC* gene.

ABSTRACT

The diversity and phylogeny of *nodC* and *nifH* genes were studied using a selection of rhizobial isolates, associated with indigenous legumes. Following restriction of nodC amplified products, 19 different nodC-genotypes were identified among 28 isolates investigated. Some of these nodC-genotypes correlated well with host species, while in other instances, symbionts, with the same nodC-genotype, were obtained from different leguminous hosts, indicating their broad host range. The RFLP method was, however, too robust and sequence analyses of representatives of these genotypes are necessary. When considering sequence variation within the *nifH* gene, the genetic distances between isolates were larger than that observed for the 16S rRNA gene. The *nifH* and 16S rDNA-based classifications were well correlated.

Keywords: *Rhizobium*, *nodC*, *nifH*, 16S rDNA classification

INTRODUCTION

The symbiotic bacteria, associated with leguminous plants, represent a diverse group of microorganisms. The taxonomical classification of these bacterial symbionts have undergone major revisions and are currently split into the genera, *Rhizobium* (Frank, 1889), *Sinorhizobium* (Chen *et al.*, 1988), *Mesorhizobium* (Jarvis *et al.*, 1997), *Allorhizobium* (de Lajudie *et al.*, 1998a), *Bradyrhizobium* (Jordan, 1982) and *Azorhizobium* (Dreyfus *et al.*, 1988). The establishment of the symbiosis is accompanied by profound developmental changes in both partners: on the roots new organs or root nodules are formed while the rhizobia inhabit these nodules as nitrogen-fixing endosymbiont bacteroids (Mergaert *et al.*, 1997).

Several classes of specific genes in rhizobia are required to establish an effective symbiosis. These include *nod* genes, responsible for the production of Nod-factors, which stimulate the host plant to produce symbiotic nodules, and *nif* genes, which produce the nitrogen fixing nitrogenase enzyme (Dixon & Wheeler, 1986). A number of genetic and biochemical mechanisms by which regulatory and structural nodulation genes control host specificity, have been identified (Roche *et al.*, 1996). Structural nodulation (*nod*) genes comprise both species-specific genes (such as *nodFEHSUZ*) and those that are common (*nodABC*) in all rhizobia (Zhang *et al.*, 2000). However, work done by Roche *et al.* (1996) suggests that *nodA* and *nodC* are also components of host-specific nodulation since NodA varies in its specificity for various fatty acid substrates, while NodC is a determinant of the length of the Nod-factor backbone. These Nod-factors, which are lipochitooligosaccharides, are the main signal molecules during the nodulation process (Perret *et al.*, 2000). Nodule development is initiated by the regulatory protein, NodD, which activates the transcription of other *nod*-genes in the presence of flavonoids, produced by the host plant (van Rhijn & Vanderleyden, 1995; Denarie *et al.*, 1996). The *nodD* genes determine the first level of host specificity since the nature and abundance of the plant flavonoids may vary according to the plant host (Schultze & Kondorosi, 1998). The common *nodABC* gene products are responsible for the formation of the core of the Nod factor, which is decorated by specific substitutions as determined by the host-specific *nod* gene products. These specific substitutions determine the plant specificity (Roche *et al.*, 1996). However, according to Perret *et al.* (2000), no

strict correlation can be drawn between the types of Nod-factors produced by rhizobia and the plants which they nodulate. As an example, *Rhizobium etli* and *Mesorhizobium loti* produce identical Nod-factors, despite them having two different host ranges (Cardenas *et al.*, 1995). Conversely, two rhizobia nodulating the same plant may produce different Nod-factors. This is the case for *R. tropici* and *R. etli*, both of which effectively nodulate the common bean plant, *Phaseolus vulgaris* (Perret *et al.*, 2000). Broad host range can also be attributed to rhizobia possessing many copies of NodD that may associate with several plant-excreted flavonoids. These flavonoid-NodD complexes are then able to activate the expression of *nod* genes. The enzymes encoded by such genes can synthesise small to large families of Nod-factors (Broughton *et al.*, 2000)

In most instances, the symbiotic genes are located on transmissible plasmids, and therefore lateral gene transfer may play an important role in the evolution of symbiosis and host range (Mergaert *et al.*, 1997). Phylogenetic studies of Nod proteins suggest that *nod* genotypes co-evolved with host plant divergence and do not correspond to 16S rRNA or *nif* phylogeny (Dobert *et al.*, 1994; Ueda *et al.*, 1995).

The phylogeny of the *nifH* gene, which codes for the Fe-protein subunit of the nitrogenase enzyme, has been reported to closely resemble that of the 16S rRNA gene, and thus these genes may share a common evolutionary history (Hennecke *et al.*, 1985; Ueda *et al.*, 1995). Instances of discordance have also been reported. Haukka *et al.* (1998) showed that bacteria with a *Sinorhizobium* chromosomal background did not necessarily share the same *nifH* genes (Eardly *et al.*, 1992).

Symbiotic performance with selected hosts is a requirement for the description of new rhizobial species (Graham *et al.*, 1991). As discussed, the genes involved in nodulation are, in most instances, located on transmissible plasmids. For this reason, phylogenies based on plant infection are considered unreliable. The study of the diversity among the symbiotic genes is still valuable to provide insight into the evolution of the *Rhizobium*-legume symbiosis. The diversity of the rhizobia, associated with indigenous leguminous plants, have been characterised by 16S rDNA sequencing and 16S-23S IGS RFLP (Chapter 5). These results indicated correspondence of the indigenous isolates to known rhizobial genera,

whereas others were more diverse and not related to the known rhizobial genera. The aim of this work was therefore to characterise the symbiotic genome of the indigenous isolates by analysing variation within the *nodC* and *nifH* genes.

MATERIALS AND METHODS

Bacterial strains

Bacterial strains used in this study are indicated in Table 6.1. Previously, Dagutat (1995) and Kruger (1998) established a culture collection of rhizobia associated with a range of legumes occurring in South Africa. Additionally, strains obtained form the root nodules of *Aspalathus linearis* were provided by the ARC-Plant Protection Research Institute (Private Bag X134, Pretoria 0001, South Africa). The 16S rRNA gene phylogeny of a selection of these isolates was determined in Chapter 5. For the purpose of this study, strains were selected to include representatives of most of the phylogenetic groupings as determined by 16S rDNA sequencing. Additionally, the *nifH* sequences of isolates 108a1 and 55 (not analysed in Chapter 5) were also included in the analyses.

Maintenance of bacterial cultures

Strains were maintained on yeast extract mannitol (YM) agar, containing (w/v): 1% mannitol, 0.5% K_2HPO_4, 0.02% $MgSO_4.7H_2O$, 0.01%, NaCl, 0.04% yeast extract and 1.5% bacteriological agar. For long term storage cultures were grown (with rigorous shaking) in yeast mannitol broth (YMB) at 28°C for approximately five days. The turbid culture broth was subsequently mixed at a 1:1 ratio with sterile 50 % (v/v) glycerol in sterile cryotubes and copies of each stored at both -20°C and -70°C.

Genomic DNA extraction

Bacterial cultures were grown in YM broth for four to seven days and genomic DNA extracted by the hexadecyltrimethyl ammonium bromide (CTAB) method as described by Wilson (1990). The integrity and concentration of the purified DNA samples was determined by agarose gel electrophoresis (Saambrook *et al.*, 1989).

PCR amplification

The PCR reactions, for both *nifh* and *nodC* amplification were carried out in a volume of 100 µl and contained 50 pmole of each primer pair, 250 µM of each dNTP, 1.5 mM $MgCl_2$, approximately 50 ng genomic DNA and 0.5 U *Taq* DNA polymerase (Southern Cross Biotechnologies). Additionally, the reaction also contained 50 mM KCl, 10 mM Tris-HCl pH
9.0 and 0.1% Triton X-100 as supplied in the reaction buffer. All amplification reactions were carried out on a Perkin Elmer GeneAmp PCR System 2400 thermocycler.

A 750 bp fragment of the *nifH* gene was amplified using universal primers nifH-F and nifH-R (Table 6.2). In most instances the *nifH* primer set, as described by Laguerre *et al.* (2001) was unable to satisfactorily amplify the *nifH* gene of the indigenous rhizobial isolates. Subsequently a new set of *nifH* primers were designed by comparing *nifH* sequence data of rhizobial reference strains with the following GenBank accession numbers: K10620 (*B. japonicum*) J01781 (*S. meliloti*), Z95228 (*Mesorhizobium* sp.), M15942 (*R. etli* bv *phaseoli*), L16503 (*Sinorhizobium* sp.) and M55226 (*R. leguminosarum* bv. *phaseoli*). The following thermal profile were found suitable for the amplification of the *nifH* gene: An initial denaturation step for 3 min at 95 °C; followed by 35 cycles of denaturation (94 °C for 30 sec), annealing (50 °C for 45 sec) and extension (72 °C for 1 min) with a final extension at 72 °C for 10 min.

For *nodC* amplification, primers nodCF, nodCFu and nodCI (Laguerre *et al.*, 2001), indicated in Table 6.2, were used. The thermal profile used for the amplification of the *nodC* gene was similar to that of the *nifH* gene, except in instances where no amplification products were obtained, the annealing temperature was adjusted to 37 °C. This led to the appearance of non-specific bands. This was resolved by cutting the desired band from the agarose, spinning it through a glasswool (Merck, South Africa) column (packed in a 0.5 ml Eppendorf tube), and the eluent, which was recovered in a 1.5 ml tube, used as template in a new PCR.

***Sequencing of nifH* DNA**

For sequencing, *nifH* PCR products were purified using Qiagen PCR Purification Kit (Southern Cross Biotechnologies, South Africa), where multiple amplification products were

observed, the desired band was gel-purified using the Geneclean™ kit (Bio 101 Inc.) according to the manufacturer's instruction. Sequencing reactions were performed on purified *nifH* PCR products using the ABI Prism BigDye™ Terminator Cycle Sequencing Ready Reaction kit (PE Applied Biosystems). Each sequencing reaction was carried out in a 5 µl volume containing approximately 100 ng template DNA, 12.5 pmole primer, and 2 µl ready reaction premix (supplied with the sequencing kit, containing the dye terminators, dNTP's, AmpliTaq DNA Polymerase, $MgCl_2$, and Tris-HCl buffer pH 9.0). The reactions were carried out in a Perkin Elmer GeneAmp PCR System 2400 thermocycler and consisted of 25 cycles of denaturation (96 °C for 10 sec), annealing (50 °C for 5 sec), and extension (60 °C for 4 min). The ClustalX programme (Thompson *et al.*, 1997) was used to analyse the nucleic acid sequences. Additional sequence data of related α-*Proteobacteria* obtained from GenBank and accession numbers are indicated in the relevant figures. The *nifH* gene sequences of the gamma-proteobacteria *Klebsiella pneumoniae* and *Azotobacter chroococcum* were also included. A distance matrix was constructed by pairwise alignment of the sequences. The phylogenetic tree was constructed from the distance matrices using the neighbour-joining method of Saitou & Nei (1987). All branch lengths were proportional to the estimated divergence along each branch. The bootstrap method (Felsenstein, 1985) was used in combination with the neighbour-joining method to estimate confidence levels of the phylogenies. The phylogenetic tree was displayed using NJplot (Perrière & Gouy, 1996).

Table 6. 2. Primers used in this study

Primer	Sequence*	Target Region	Reference
NifH-F	5' CGGGAAGGGCGGAATCGGCAAG 3'	*nifH*	This work
NifH-R	5' GCATGTCCTCGAGCTC(AT)TCCAT 3'	*nifH*	This work
nodCF	5' AYGTHGTYGAYGACGGTTC 3'	*nodC*	Laguerre *et al.* (2001)
nodCFu	5' AYGTHGTYGAYGACGGITC 3'	*nodC*	Laguerre *et al.* (2001)
nodCI	5' CGYGACAGCCANTCKCTATTG	*nodC*	Laguerre *et al.* (2001)

*Abbreviations: Y = C or T H = A, C or T I = inosine N = A, C, G or T K = G or T.

nodC RFLP analysis

The *nodC*-amplified products were used directly in a restriction reaction, without any prior purification. The digestion reactions were performed using 10 U of each of the following restriction enzymes *Hae*III, *Cfo*I, *Hinf*I, *Msp*I and *Rsa*I (Roche Molecular Diagnostics, South Africa) in optimal buffers as prescribed by the supplier. Electrophoresis of the restricted DNA was carried out on a 3% agarose gel in a Hybaid Maxi Gel System at 80 V for 90 min. DNA molecular weight marker VIII (Roche Molecular Diagnostics, South Africa) was included at multiple positions on gels as a standard. Restriction patterns were identified visually for each of the restriction enzymes used. A matrix, describing the presence (indicated by 1) or absence (indicated by zero), of a particular pattern within an isolate was generated. In this way a specific profile for each isolate was compiled. The Dice coefficient (Nei & Li, 1979) within the Bionum computer programme (Applied Maths, Kortrijk, Belgium) was used to construct a distance matrix from this one (1)-zero (0) profiles. These distance values were subsequently analysed to generate a tree using the UPGMA algorithm in GelCompar 4.0 (Applied Maths, Kortrijk, Belgium). Restrictions patterns of the *Mesorhizobium* and *Sinorhizobium* were obtained from Laguerre *et al.* (2001), while the RFLP patterns of the *Bradyrhizobium* spp. were determined in this study. Other rhizobial genera were not considered in the final dendrogram construction since initial inspection of the restriction pattern showed no correspondence to such genera.

RESULTS and DISCUSSION

The collection of isolates investigated here was obtained from root nodules from a range of diverse leguminous plants. However, the collection comprised isolates mainly from host plants of the Papilionoideae subfamily, with the Mimosoideae and the Caesalpinioideae represented by one and two isolates, respectively. In general each of the isolates was obtained from a different host plant. Isolates, obtained from the same host plant were the eight symbionts obtained from *Aspalathus linearis*, isolates 70a and 70b2 from *Crotalaria brachycarpa* and 82a, 48a and 48b from *Tephrosia purpurea*.

PCR amplification of the *nodC* and *nifH* gene fragments

A 900 bp fragment of *nodC* gene, which determines the length of the Nod-factor backbone, was obtained for most isolates. Isolates, SA3 and 66a1t1, repeatedly gave an amplification

product of approximately 1100 bp, irrespective of the forward primer used. The nodCFu-nodCI primer pair was successfully used in most of the amplification reactions. For some isolates, as indicated in Table 6.1, no *nodC*-amplification product could be obtained. In an effort to find a basis for this lack of amplification, their phylogenetic positions based on the conserved 16S rRNA gene and the variable 16S-23S intergenic spacer region (IGS), as described in Chapter 5 (Fig. 5.1 & Fig. 5.4) were considered. Based on IGS phylogeny, isolates 103b, 3a, 13b, 15b, and xhj13 were divergent genotypes and not closely related to any of the reference species of the 16S rDNA subgroups to which they belonged. Laguerre *et al.* (2001) reported that *nodC* primers, similar to those used in this study, were unable to amplify the *nodC* fragment of *Sinorhizobium saheli* bv. acaciae. The 16S rDNA sequences of isolates xhj12FR and UP27b were closely related to *Sinorhizobium meliloti* and a *nodC* fragment for these isolates could not be obtained as well. It therefore appears that the current primers are not sufficient to cover the heterogeneity associated with the *nodC* gene. For the other strains, which also failed to amplify (as indicated in Table 6.1) no obvious causes were evident for this lack of visible amplified products.

A 750 bp *nifH* fragment was obtained for most isolates, although some showed non-specific bands smaller than the expected size fragment.

Table 6.1. Isolates investigated in this study and a summary of RFLP data following digestion of *nodC*-amplified products with indicated enzymes.

Alphabetical denotations refer to the first letter of the relevant enzyme. Distinctive patterns are indicated by numerical designations. Strains, where no amplified product was obtained indicated by (nd). The *Hinf*I pattern, Hi2, refers to no restriction site. Subfamilies of the Fabaceae indicated in brackets: P: Papilionoideae, M: Mimosoideae, C: Caesalpinioideae

Isolate number	Host legume/ Reference strain	*nodC* restriction patterns after digestion with				
		***Hae*III**	***Hinf*I**	***Cfo*I**	***Msp*I**	***Rsa*I**
2	*Rhynchosia nervosa* (P)	H3	Hi2	C2	M2	R2
29d	*Alysicarpus rugosus* (P)	H3	Hi2	C2	M2	R2
42bs	*Neonotonia wightii* (P)	H1	Hi1	C1	M1	R1
70a	*Crotalaria brachycarpa* (P)	H3	Hi2	C2	M2	R2
82a	*Tephrosia purpurea* (P)	H3	Hi2	C2	M2	R2
103b	*Indigofera hilaris* (P)	nd	nd	nd	nd	nd
40a	*Sesbania bispinosa* (P)	H2	Hi2	C3	M3	R3
48a	*Tephrosia purpurea*(P)	H2	Hi2	C3	M3	R3
48b	*Tephrosia purpurea* (P)	H2	Hi2	C3	M3	R3
49b	*Indigofera melanadenia* (P)	H2	Hi2	C3	M3	R3
33b	*Crotalaria* sp. (P)	nd	nd	nd	nd	nd
123b	*Desmodium* repandum (P)	H2	Hi2	C3	M4	R4
114d	*Pseudarthria hookeri* (P)	H4	Hi3	C4	M5	R4
46c2	*Acacia sieberana* var *woodii* (M)	H5	Hi4	C5	M4	R5
125e	*Indigofera woodii* (P)	H6	Hi2	C3	M4	R3
70b2	*Crotalaria brachycarpa* (P)	H2	Hi2	C6	M6	R3
104a1	*Chamaecrista* sp. (C)	H2	Hi2	C6	M6	R3
3a	*Trifolium* sp.(P)	nd	nd	nd	nd	nd
13b	*Lotononis bainesii* (P)	nd	nd	nd	nd	nd
15b	*Desmodium tortuosum* (P)	nd	nd	nd	nd	nd
LMG 18230	*Bradyrhizobium liaoningense*	H2	Hi3	C6	M7	R6
LMG 6138	*Bradyrhizobium japonicum*	H2	Hi3	C6	M7	R6
LMG 6134	*Bradyrhizobium elkanii*	H10	H5	C3	M8	R6
xhj15	*Aspalathus linearis* (P)	nd	nd	nd	nd	nd
xhj13	*Aspalathus linearis* (P)	nd	nd	nd	nd	nd
xhm5	*Aspalathus cordata* (P)	nd	nd	nd	nd	nd
xhj18	*Aspalathus linearis* (P)	H6	Hi2	C7	M9	R7
xhj23	*Aspalathus linearis* (P)	H6	Hi2	C7	M9	R7
xhj7	*Aspalathus linearis* (P)	H6	Hi2	C8	M10	R8
xhj8	*Aspalathus linearis* (P)	H6	Hi2	C8	M10	R8
xhj20	*Aspalathus linearis* (P)	H6	Hi2	C7	M10	R7
xhj27	*Aspalathus linearis* (P)	H7	Hi2	C8	M9	R14
xhj26	*Aspalathus linearis* (P)	H6	Hi2	C8	M10	R8
xhj12FR	*Aspalathus linearis* (P)	nd	nd	nd	nd	nd
98d2	*Bolusanthus speciosus* (P)	nd	nd	nd	nd	nd
128a	*Melolobium obcordatum* (P)	H1	Hi2	C8	M12	R7
MSDJ 2184	*Mesorhizobium* sp.	H7	Hi7	C9	M13	R9
ACCC19665	*Mesorhizobium amorphae*	H6	Hi8	C10	M14	R10
7653R	*Mesorhizobium huakuii*	H1	Hi9	C11	M15	R11
USDA3233	*Mesorhizobium ciceri*	H8	H10	C12	M16	R12
xhj12s	Aspalathus *linearis* (P)	H8	Hi6	C6	M11	R14
USDA1002	*Sinorhizobium meliloti*	H9	Hi6	C13	M17	R13
M1	*Sinorhizobium medicae*	H10	Hi11	C14	M18	R14
36b	*Neonotonia wightii* (P)	H2	Hi12	C6	M6	R15
UP27b	*Crotalaria* damrensis (P)	nd	nd	nd	nd	nd
SA3	*Trifolium africanum* (P)	H11	Hi13	C15	M19	R16
UP26b	*Crotalaria* damrensis (P)	H12	Hi14	C16	M20	R17
66a1t1	*Acacia robusta* (M)	H13	Hi2	C7	M21	R17
7b	*Alysicarpus rugosus* (P)	nd	nd	nd	nd	nd
60	*Rhynchosia monophylla* (P)	nd	nd	nd	nd	nd
PL10a	*Argyrolobium tomentosum* (P)	nd	nd	nd	nd	nd

Diversity of *nodC* genotypes and host specificity

The different restrictions patterns obtained after digestion with five restriction enzymes are indicated in Table 6.1. Among the 28 isolates analysed, 19 different *nodC* genotypes were present. This correlated well with results obtained by Laguerre *et al.* (2001), who also found 45 nod types among 82 strains investigated. These authors also concluded that the *nodC* gene was highly polymorphic. According to Mergaert *et al.* (1997) the *nodABC* genes were previously thought to be functionally conserved in all rhizobia. However, it has been demonstrated that different alleles of the *nodA* and *nodC* genes are present in the rhizobial gene pool and that this allelic variation contributes to host range determination. The inability of the current primers to generate an amplification product in some instances might also be related to the heterogeneity of the *nodC* gene. Of the different *nodC* types present among the selected indigenous isolates, none showed 100% correspondence to the included rhizobial reference strains.

The different restriction patterns were used to construct a dendrogram (Fig. 6.1), which could be divided into two sections. Within section 1, different clusters were present, some of which (clusters A and B) were clearly host species-specific clusters. The *nodC* genotype of isolates associated with *Aspalathus linearis* was found to be polymorphic since two related clusters (A and B) were observed, while xhj27 only had a *Hinf*I restriction pattern in common with members of clusters A and B and xhj12s was distantly related to *S. meliloti* (section 2). According to Dobert *et al.* (1994) and Ueda *et al.* (1995) different *nod* genotypes that are found in rhizobia have evolved to allow optimal nodulation of the corresponding host legumes.

Members of clusters C, E and F individually displayed identical *nodC* types despite the fact that they were obtained from different host plants. A previous investigation (Chapter 5) showed similar clustering of these isolates into highly homologous groups based on 16S rDNA sequences. Subsequent analysis of their 16S-23S intergenic spacer (IGS) regions, which have higher discriminatory ability (Gurtler *et al.*, 1996), also showed that the isolates within these groups are closely related. In the absence of DNA homology data, which is an important criterion for species delineation, it is difficult to speculate whether strains in these groups are the same species or not. However, according to Willems *et al.* (2001c), IGS

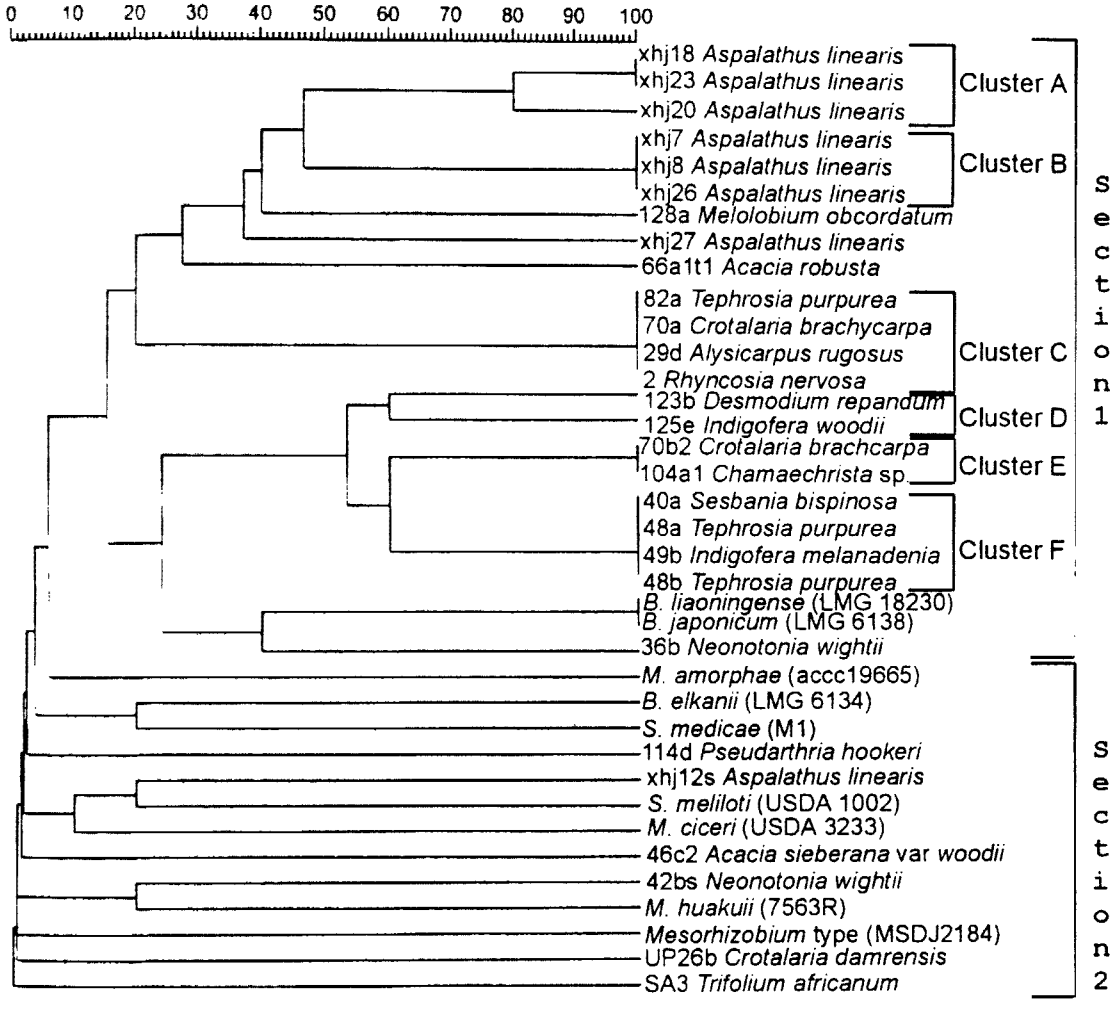

Figure 6.1. Dendrogram based on the combined *Hae*III, *Hin*fI, *Cfo*I, *Msp*I and *Rsa*I restriction patterns of *nodC* amplified product. The host species from which these isolates were obtained are shown in italics. Information of the restriction patterns of the genera *Mesorhizobium* and *Sinorhizobium* were obtained from Laguerre *et al.* (2001). The *x*-axis represents the correlation between strains.

variability sometimes correlates well with DNA homology data for the *Bradyrhizobium* genus. Preliminary chromosomal evidence (16S rDNA sequence and IGS) therefore suggests that the respective isolates in clusters C, E and F, which share similar *nodC* types, may represent three species. In the light of this evidence, it would appear that these strains have a broad host range since they were isolated from different host legumes.

The *nodC* types of isolates 36b, 114d, 46c2, 42bs, UP26b and SA3 showed no close relation to any of the other nod types obtained in this study. Laguerre *et al.* (2001) concluded that the RFLP method was insufficient to determine the true phylogenetic relationships of the *nodC* gene since few restriction sites, within this genomic region, were conserved among species or biovars. In an effort to determine the true phylogenetic relationships between these *nodC* types, nucleotide sequencing of representatives of such types will be necessary.

Phylogeny based on partial *nifH* gene sequences and comparison with 16S rDNA classification

Partial *nifH* sequence data were used to construct a phylogeny of a selection of the indigenous strains. The *nifH* sequence of rhizobial reference strains: *B. liaoningense* (LMG 18230), *M. loti* (LMG 6125), *M. amorphae* (LMG 18977), *M. plurifarium* (LMG 11892), *S. xinjiangensis* (LMG 17930) were also determined in this study since they were not available in the GenBank sequence database. In general, the constructed phylogenetic tree (Fig. 6.2) correlated well with the 16S rRNA phylogeny (Chapter 5) in accordance with previous reports (Hennecke *et* al., 1985; Ueda *et al.*, 1995). The genetic distances, based on *nifH* data, were larger than that observed among 16S rDNA sequences. As an example, isolate 70b2 showed almost 99% sequence homology with *B. japonicum*, while it showed only 83% sequence identity based on *nifH* sequence data. 16S rDNA sequences also showed that isolates such as 40a, 48a, 49b and 60, were divergent bradyrhizobial strains. The *nifH* phylogeny confirms their divergent nature because, although they were present in the same cluster as *B. japonicum* and *B. liaoningense*, they were carried on a separate lineage. The *nifH* sequences of isolates from *Aspalathus linearis*, which showed highest 16S rDNA homology to *M. loti*, were compared to *M. loti nifH* sequences from two different sources: the *M. loti* (LMG 6125) *nifH* sequence, as determined in this study and *M. loti* genome

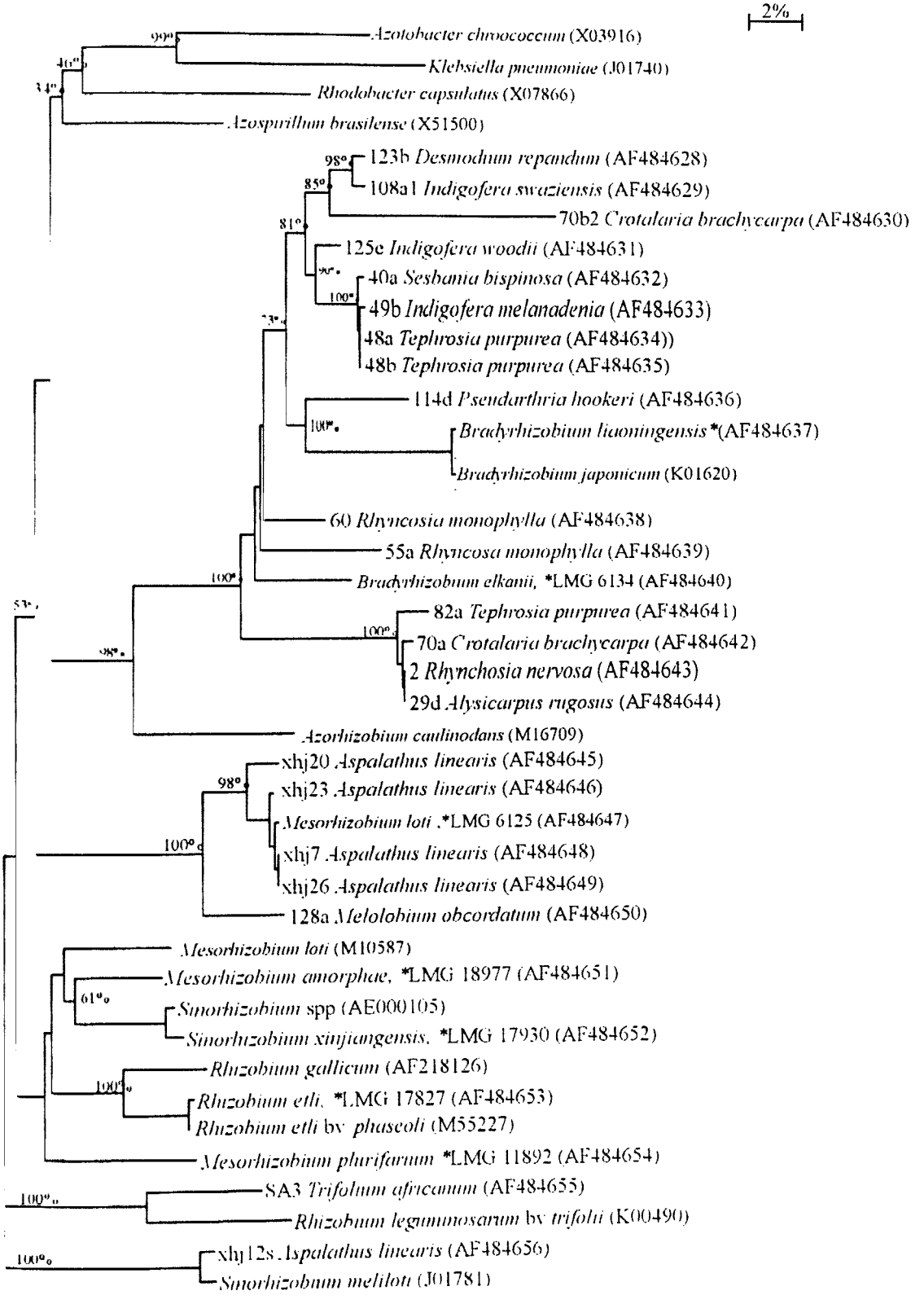

Figure 6.2. Phylogenetic tree, based on a 460 bp fragment of the *nifH* gene, expressing the relationship of a selection of the indigenous rhizobia to other rhizobial genera and other nitrogen fixing strains of the alpha- and gamma-proteobacteria. The tree was generated using the Neighbour-joining method of Saitou & Nei (1987). Horisontal branches are equal to the phylogenetic distances of which the scale indicate 2%. Bootstrap values of some of the major branching points are indicated. GenBank accession numbers are given in brackets. The *nifH* sequences of reference strains (indicated by *) were not available and were determined in this study.

sequencing project (GenBank accession number M10587). A discrepancy was observed between these two sequences, since they had different positions on the phylogenetic tree, with the *Aspalathus linearis* isolates being closer related to LMG 6125.

When Haukka *et al.* (1998) investigated the diversity of a group of strains from leguminous trees growing in Africa and Latin America, *Sinorhizobium* strains with different *nifH* sequence were found. In this case the *nifH* tree was not consistent with the tree based on 16S rDNA sequences, suggesting the occurrence of lateral gene transfer. Among the indigenous isolates investigated here, no evidence of lateral gene transfer could be indicated.

Finally, the results presented here showed that the bacterial symbionts, forming symbiotic associations with the indigenous leguminous plants, harbour a diverse range of *nodC* genotypes. The extent of this diversity will be reflected more accurately once sequence analyses are performed on representatives of such genotypes and when more isolates are investigated from a common host species. The *nifH* phylogeny reflects the 16S rRNA gene phylogeny. However, it is unlikely that this would be true for symbionts from all leguminous plants.