Dissertation

Submitted to the
Combined Faculties for the Natural Sciences and for Mathematics
of the Ruperto-Carola University of Heidelberg, Germany

for the degree of
Doctor of Natural Sciences

Presented by

Md. Harun-or Rashid
born in Sherpur, Bangladesh

Date of oral examination:

Genetic diversity of rhizobia nodulating lentils (*Lens culinaris* Medik.)

Referees: Prof. Dr. Michael Wink

Prof. Dr. Thomas Rausch

Acknowledgements

This thesis would not have been possible without the help of many people who supported me and my work during my study period.

I express my deepest gratitude to my supervisors Prof. Michael Wink and Prof. Thomas Rausch for accepting me as Ph.D. student and for providing all necessary support to finish my research project. I owe my deepest gratitude to both of them for their critical advice and enormous support during my study period.

I am grateful to Prof. Michael Wink for freedom in research, scientific guidance, confidence and encouragement. It was a great pleasure for me to work on rhizobia project in his laboratory.

I am grateful to Prof. Thomas Rausch for his inspirations, critical suggestions for research and general administration, and for providing research facility at glasshouse condition.

I am grateful to my wife, Afrid Akter for her limitless inspiration, understanding, sacrifice and taking care of my children in Bangladesh during my study at Heidelberg University.

I am deeply indebted to the EMMA authority for providing a scholarship to me for doing my Ph.D. at Heidelberg University. I appreciate the help from the Bangladesh Agricultural Research Council, Bangladesh for providing partial funding for the field work in Bangladesh and the DAAD-STIBET program authority for providing a thesis completion grant to me.

I would like to thank Prof. J.P.W. Young, University of York (UK) for providing reference strains and critical suggestions to my work. I am grateful to Dr. Holger Schäfer for his nice cooperations regarding collections of primers, chemicals, soil samples and fruitfull discussions during my study period. I would like to thank Dr. Javier Gonzalez for his critical comments and suggestions regarding phylogenetic interpretations and software use. Special thanks to Markus Santhosh Braun for his nice cooperation for solving many problems related to German language and general administration during laboratory work, summary writing and daily life.

I would like to show my gratitude to Prof. M. A. Hashem, Bangladesh Agricultural University and Dr. M. A. Satter, Director General, Bangladesh Institute of Nuclear

Agriculture for encouraging me to do my research project in Germany. I am grateful to my friend Dr. Martin Krehenbrink, Chief technical officer, CYSAL, Germany, for improving language and fruitful discussion while writing up my thesis.

I would like to show my gratitude to Polash Sarker, Agriculture Officer, Dhaka, Nazim Uddin Shakh, Deputy Director, BADC, Earshad Ali, SSAE, Rajshahi, Golam Rasul, ASO and Shahin Akther, ASO, BINA, for their help to collect field grown lentil nodules from Bangladesh.

Several people have contributed to collect soil samples from Germany, Turkey and Syria. I am grateful to all of you: Heidi Staudter, Annika Heinemann, Vanessa Erbe, Johannes Reiner and Beate Waibel from Germany, Razan Hamoud, Shirin Hamoud and Rasha Abou Aleinein from Syria ,and Dr. Tamer Albayrak from Turkey for your important contribution for making a good story about my research project. I am grateful to Dr. Sabine Grube, Hoehenheim Unversity, Mr. Woldae Mammel, president (Vorsitzender) of the Friends Association (Förderverein), Alblinsen and Mr. Markus Santhosh Braun for their help to collect field grown lentil nodules from Germany.

I am grateful to Heidi Staudter and Hedwig Sauer-Gürth for their lot of help and technical assistance to my laboratory work. Thank to Astrid Backhaus and Dieter Holzmann for their general help to my laboratory work. Special thank to Michael Schilbach at COS for his technical assistance during my work at glasshouse conditions. I would like to thank Mr. Philipp Kremer for helping in the mapping of sampling localities. I am indebted to my colleagues to support me during my study period. I would like to thank our secretary Petra Fellhauer for her lots of help and nice cooperations during my study period.

This work has been performed according to Bangladesh and German law.

Md. Harun-or Rashid

Table of contents

Summary

Lentil is not only the oldest legume crop but also the oldest of the crops that have been domesticated in the Fertile Crescent and distributed to other regions during the Bronze Age, making it an ideal model to study the evolution of rhizobia associated with crop legumes. This study investigates lentil-nodulating rhizobia from the region where lentil originated (Turkey and Syria) and from regions to which lentil was introduced later (Germany and Bangladesh). There are few studies on lentil-nodulating rhizobia, and no phylogenetic studies on lentil rhizobia using multi locus sequence analyses. Therefore, rhizobia from lentil nodules were chosen to study 1) the genetic diversity 2) the taxonomic position and 3) the transmissible nature of nodulation genes. I have sequenced four housekeeping genes (16S rRNA, *recA*, *atpD*, *glnII*) and three nodulation genes (*nodA*, *nodC*, *nodD*) and analyzed these using phylogenetic and population genetic approaches to achieve these objectives. To supplement these approaches I have also used DNA fingerprinting and phenotypic characterization. Moreover, the symbiotic performance was assessed by nodulation and cross inoculation tests. I identified four different lineages of rhizobia associated with lentil, of which three are new and endemic to Bangladesh, and one lineage was found in the Mediterranean region and Central Europe. The new lineages from Bangladesh are close to *Rhizobium etli* and correspond to new species in the genus *Rhizobium*. The endemic lentil grex *pilosae* may have played a significant role in the origin of these new lineages in Bangladesh. The single lineage from the Mediterranean and Central Europe belongs to *Rhizobium leguminosarum.* The association of *Rhizobium leguminosarum* with lentil at the centre of lentil origin and in countries where lentil was introduced later suggests that *Rhizobium leguminosarum* is the original symbiont of lentil. Lentil seeds might have played a significant role in the initial dispersal of *Rhizobium leguminosarum* within the Middle East and on to other countries. Analysis of nodulation genes showed that they are prone to horizontal transfer between different chromosomal lineages and sub-lineages of rhizobia. Nodulation genes showed bias to their geographical origin, evidencing that plasmid-borne characters in bacteria rapidly change according to their adaptation to particular environment.

Key words: *Rhizobium*, *Lens culinaris*, nodulation, multilocus analysis, fingerprint, phylogeny

Zusammenfassung

Die Linse ist die älteste Hülsenfrucht und zugleich die älteste der im Fruchtbaren Halbmond domestizierten Kulturpflanzen. Im Bronzezeitalter wurde sie in andere Regionen eingeführt, was sie zu einem idealen Studienobjekt der Evolution von mit Kulturpflanzen assoziierten Rhizobien macht. Diese Arbeit untersucht knöllchenbildende Rhizobien der Linse aus den Ursprungsländern (Türkei und Syrien) und aus Gebieten, in die sie erst später eingeführt wurde (Deutschland und Bangladesch). Bislang gibt es nur wenige Studien an Knöllchenbakterien von Linsen; Untersuchungen, phylogenetische Analysen mit Hilfe des Multi Locus Sequencing existieren nicht. Aus diesem Grund wurden aus Linsenknöllchen stammende Rhizobien gewählt, um 1) deren genetische Vielfalt, 2) ihre taxonomische Position und 3) die Übertragbarkeit von Nodulationsgenen zu untersuchen. Um diese Ziele zu erreichen, habe ich vier Haushaltsgene (16S rRNA, *recA*, *atpD, glnII*) und drei Nodulationsgene (*nodA, nodC, nodD*) sequenziert und diese mittels phylogenetischer und populationsgenetischer Methoden analysiert. Ergänzend habe ich DNA-Fingerprints erstellt und die Rhizobien phänotypisch charakterisiert. Zusätzlich wurde die symbiotische Effizienz durch Nodulationstests und Kreuzbeimpfungen bestimmt. Ich habe vier unterschiedliche Abstammungslinien von Linsen-Rhizobien identifiziert. Drei davon waren bislang unbekannt und sind in Bangladesch endemisch. Die verbleibende kommt im Mittelmeergebiet und Mitteleuropa vor. Die neuen Abstammungslinien aus Bangladesch sind nahe Verwandte von *Rhizobium etli* und bilden innerhalb der Gattung *Rhizobium* drei neue Arten. Die endemische Linsengrex *pilosae* könnte eine entscheidende Rolle hinsichtlich des Ursprungs der neuen Abstammungslinien Bangladeschs gespielt haben. Die Abstammungslinie der Mittelmeergebiete und Mitteleuropas gehört zu *Rhizobium leguminosarum*. Das Vorkommen von *Rhizobium leguminosarum* an Linsen im Kern des Ursprungsgebiets der Pflanze und in Ländern, in die die Linse später eingeführt wurde, lässt vermuten, dass *Rhizobium leguminosarum* der ursprüngliche Symbiont dieser Kulturpflanze ist. Linsensamen könnten für die Verbreitung von *Rhizobium leguminosarum* in den Nahen Osten und andere Länder entscheidend gewesen sein. Die Analyse der Nodulationsgene hat gezeigt, dass leicht ein horizontaler Gentransfer zwischen unterschiedlichen chromosomalen- und Sub-Abstammungslinien von Rhizobien stattfinden kann. Die Variabilität der Nodulationsgene wies eine Korrelation mit deren geographischen Ursprung auf, was zeigt, dass sich durch Plasmide verliehene Eigenschaften von Bakterien schnell an ihre spezielle Umgebung anpassen.

This thesis is based on the following manuscripts:

1. **Rashid MH,** Schafer H, Gonzalez H & Wink M (2012) Genetic diversity of rhizobia nodulating lentil (*Lens culinaris*) in Bangladesh. *Syst Appl Microbiol* **35**: 98 – 109.

2. **Rashid MH,** Gonzalez H, Young JPW and Wink M (2013) *Rhizobium leguminosarum* symbiovar *viciae* is the symbiont of lentils in the Middle East and Europe but not in Bangladesh. Submitted to the "*FEMS Microbiology Ecology*" (under review).

1. Introduction

1.1 Rhizobia

Nitrogen is an essential nutrient for all living organisms and necessary for the production of high-yield and high-quality agricultural crops. Although molecular nitrogen (N_2) is the most abundant gas in the atmosphere, it is unavailable to plants in its elemental form. Rhizobia are a group of bacteria that have the capacity to form nodules on legume roots (and occasionally on stems) and can fix atmospheric nitrogen to partially or fully meet the nitrogen requirements of the host plant. To describe bacteria from root nodules, Frank (1889) proposed name "rhizobia", and after this proposal all nodule-forming bacteria have been known as rhizobia. Biological nitrogen fixation (BNF: atmospheric nitrogen fixation through different members of prokaryotes, specifically by diazotrophs) contributes approximately 16% of total nitrogen input in crop land (Ollivier *et al.*, 2011). Rhizobia are a major contributor to BNF, and the legume-rhizobium symbiosis can fix up to 450 Kg N/ha/year (Unkovich & Pate, 2000). Rhizobia as a group are not monophyletic and have been classifed as α and β rhizobia (Moulin *et al.*, 2002). Already 163 species from 12 genera (http://edzna.ccg.unam.mx/rhizobial-taxonomy) have been described. However, further study of the genetic diversity of rhizobia helps to understand the evolutionary histories of the legume-rhizobium symbiosis and helps to devise effective planning strategies to achieve the maximum benefit from legume-rhizobium symbiosis.

1.2 Legume-Rhizobium symbiosis

Legumes are the third largest flowering plant family, containing 700 genera and about 20,000 species. It is an extremely diverse plant family with a great number of different characters, and a few characters are common to all species. Nodulation is an important characteristic of majority legumes but it is absent in the earliest divergent lineages of this family. Legume evolved about 65 MYA and nodulation have evolved in legumes after the origin of this family approximately 30 MY (Lavin *et al.*, 2005; Bruneau *et al.*, 2008; Doyle, 2011) later.

Nodulation is a complex issue and there is no simple explanation for its origin (Doyle, 1998; 2003; Sprent, 2001). Nodulation takes place in the legume sub-family Papilionoideae and Mimosoideae, and in the tribe Caesalpinieae and core Cassieae. Most of the members of the Papilionoideae and about 90% of the Mimosoid members are thought to be nodulated by rhizobia for nitrogen fixation (Doyle, 2011). Nodulating or

non-nodulating legumes have a high demand for nitrogen, largely in the leaf where photosynthesis or the accumulation of nitrogen-rich defensive compounds occurrs. Hence, it has been concluded that legumes owe their evolutionary success to nodulation and subsequent access to atmospheric nitrogen (McKey, 1994).

Rhizobia produce nodulation (nod) factors after a specific interaction with their host plant, and thus it has been assumed that rhizobia have coevolved with their host plants (Perret *et al.*, 2000). The strong correlation between host plant and nodulation genotypes denotes the importance of the host on nodulation genotypes (Young & Wexler, 1988; Laguerre *et al.*, 1992; 1993; Black *et al.*, 2012). However, from two model legumes, *Lotus japonicus* and *Medicago truncatula,* 26 genes have been cloned which are involved in recognition of rhizobia, the nodulation signal cascade, infection, the nodulation process and the regulation of nitrogen fixation (Kouchi *et al.*, 2010; and reference therein). However, host association also is important for shaping the genetic divergence of nodulation and housekeeping genes in rhizobia (Wernegreen & Riley, 1999).

1.3 Lentil cultivation status in different countries

Root-nodulating bacteria were first described from different members of the legume tribe *Vicieae*. The tribe *Vicieae* is composed of several closely-related genera with many intermediate forms. Therefore, this tribe contains several newly proposed genera. However, there are four well-recognized genera such as *Vicia*, *Lathyrus*, *Pisium* and *Lens* in the tribe *Vicieae*.

Lentil (*Lens culinaris*) is an important and popular legume employed for human and animal nutrition, and soil fertility management. *Lens* is a Latin word, which describes the shape of cultivated lentil seeds. The word and seed are very distinct from other legumes, and it is therefore easy to identify lentil from historical texts and archeological sites (Cubero, 1981; and references therein). Lentil were domesticated in the Fertile Crescent around 9,000 years ago (Toklu *et al.*, 2009; and reference therein) and remains an important seed legume crop cultivated worldwide (Sarker & Erskine, 2006). The region for the domestication of lentl consists of Southeastern Turkey and Northern Syria, including the sources of the Tigris and the Euphrates rivers (Lev-Yadun *et al.*, 2000). *Lens culinaris* is the putative ancestor of domesticated lentil (Barulina, 1930; Cubero, 1981).The taxonomy of cultivated lentil subspecies and grex is shown in table 1 (after Cubero, 1981):

Table 1. Taxonomy of cultivated lentil

Species	Subspecies	Race	Grex
Lens culinaris	*nigricans*	*macrosperma*	*europeae*
		microsperma	*asiaticae*
			intermediae
	orientalis		*subspontaneae*
			aethiopicae
			pilosae
	culinaris		

After domestication in the cradle of agriculture, lentil spread to Cyprus in the Neolithic (Erskine *et al.*, 1994). Lentil disseminated from Southeastern Europe to Central Europe around the 5,000 years BC via the Danube. From Europe, it dispersed to the Nile Valley and from there to Ethiopia. However, in Georgia, lentil was propagated during the 5,000 and 4,000 years BC and in the Indian sub-continent around 2,500 – 2,000 years BC (Sonnante *et al.*, 2009; and references therein). Lentil is now grown all over the world (Indian sub-continent, Middle East, North Africa, South Europe, the North and South America and Australia, Chahota *et al.*, 2007).

Lentil cultivation in Bangladesh

Legumes play an important role in agriculture and daily diet in Bangladesh. Around 5.2% of cultivable lands are subject to legume cultivation (Rahman *et al.*, 2009). Because poor people cannot afford expensive fish and meat, pulses have been known for a long time as the “meat of the poor people” in Bangladesh. Although legumes have been grown in Bangladesh for a long time, farmers have largely been cultivating pulse legumes without applying major agricultural inputs like fertilizers, irrigation and plant protection. A number of pulse crops like vetch *(Lathyrus sativus* L.), lentil *(Lens culinaris,* Medic), chickpea *(Cicer arietinum* L.), black gram *(Vigna mungo* L.), mung bean *(Vigna radiata* L.), and cowpea *(Vigna unguiculata* L.) are grown in Bangladesh. Among these crops, lentil is the most popular and has been cultivated since ancient times. In 2010, lentil was grown on 9,199 hectares of land and the total production was 71,100 tonnes (FAOSTAT-Agriculture, 2010). The grex *pilosae* in the race microsperma of *Lens culinaris* is the only

cultivated lentil in Bangladesh (Dr. A. Sarker, Dr. A. Rahman and Dr. M. A. Samad, personal communication).

Lentil cultivation in Germany

Germany has a history of cultivation of different pulse crops like dry bean, horse bean, lupin, pea and lentil. In 2010, about 103,800 hectares of land were used for different pulse crop production in Germany. Of the area cultivated with legumes (0.3% of the total cultivated area, http://www.aid.de) about 17,200 hectares were used for pea production (FAOSTAT-Agriculture, 2010). In 2012 about 200 hectare of land was under lentil cultivation in Germany (Dr. Sabine Gruber; Mr. Woldemar Mammel, personal communication). For several decades, little attention was paid to lentil cultivation and research in Germany. Nowadays, there is an increasing market for high-priced lentil in Germany. Considering the various advantages like nitrogen fixation potential, high nutritive values and essentiality for mix cropping, more and more farmers are beginning to reintroduce lentil to German organic farming, and scientists are paying more attention to lentil research (Wang *et al.*, 2012).

Lentil cultivation in Syria

Different pulse legumes like bean, broad bean, chickpea, lentils, lupin, pea and vetch are grown in Syria (FAOSTAT-Agriculture, 2010). Lentil is an important pulse crop in Syria. It does not only provide cheap dietary protein but also plays significant role for farm land and farmers. Moreover, lentil cultivation is an integral part of crop rotation to maintain soil fertility and as sures that farmers are getting a higher more economic return from cultivation. Among the different provinces, Aleppo, Idlib, Al Hasakah and Hama are the main (80%) lentil producers in Syria, though Damascus and As Suwayda also contribute (20%) to the national production (www.houseoflentils.com). In 2010, about 242,620 hectares were used for legume production and 131,100 hectares were used for lentil cultivation (FAOSTAT-Agriculture, 2010).

Lentil cultivation in Turkey

Since ancient times, pulses have been cultivated and consumed in large quantities in Turkey. Turkey is also the third largest lentil producing country in the world after India and Canada (FAOSTAT-Agriculture, 2007). Major pulses like lentil, chickpea, white bean, red bean, broad bean and vetch have been grown in Turkey

(www.economy.gov.tr; FAOSTAT-Agriculture, 2010), but pulse production is significantly influenced by lentil production (http://www.invest.gov.tr/en). About 895,689 hectares of land are used for the production of different pulses and 234,378 hectares were used for lentil cultivation in 2010 (FAOSTAT-Agriculture, 2010). Although pulse cultivation is distributed throughout Turkey, some regions are important for specific pulse crops. For example, red lentil is grown in the Southeast Anatolia, green lentil, chickpea and dry bean are grown in central Anatolia, and broad bean and dry pea are cultivated in the Aegean and Marmara regions (www.economy.gov.tr). Although *macrosperma* is the dominating race in Turkey, the *microsperma* race is also available in some regions like South Turkey (Toklu *et al.*, 2009).

1.4 Taxonomy of bacteria

Systematics is the branch of biology which deals with the classification of living organisms by describing their diversity and relationships to one another. Systematics has three major divisions; taxonomy, classification and phylogenetics (Wink, 2007).

A group of organism which is specifically defined on the basis of certain characters is known as a taxon, and the description and naming of taxa is known as taxonomy. Taxonomy organizes different taxa on the basis of their similarity and is therefore important for the identification of organisms at field level. However, taxa may not always reflect the evolutionary relationship of organisms. As for eukaryotes there are different levels in bacterial taxonomy, although kingdoms have not been proposed for bacteria. In bacterial taxonomy the topmost level is the domain, which is followed by phylum, family, genus, species and sub-species (Brenner *et al.*, 2005). However, taxonomy is not fixed, but rather always prone to change with the development of new ideas. To implement a new idea it is necessary to change the taxonomy, and subsequently changes have occurred in classification, nomenclature and identification.

There are three major concepts for defining species, such as the typological /morphological concept, the biological species concept, and the phylogenetic /evolutionary species concept. According to the morphological species concept, a species combines a group of organisms based on sufficiently close, shared and fixed morphological or anatomical properties. In the biological species concept (Mayr, 1963), a species is a group of populations whose members can potentially inbreed and are reproductively isolated from others group.

The phylogenetic species is the most recent concept. In this concept, a species is a group of organisms that share a common ancestor and can be separated from other species by distinctive characters. The evolutionary species concept describes that maintain integrity from other lineages through both time and space. During evolution, members of the lineage can diverge and become an independent species (Wink, 2007 ; and references therein).

Phylogenetics is the field of biology that deals with identifying and understanding the evolutionary relationships between different kinds of life on Earth, and is the basis for evolutionary systematics. Classification is the organization of information about diversity, and arranges organisms into a c onvenient hierarchical system such as the Linnean system.

1.5 Rhizobial taxonomy

From the very begening of bacterial taxonomy, rhizobial taxonomy was significantly influenced by the nodulating host. In the initial classification of bacteria (Bergey *et al.*, 1923), rhizobia were described as Gram-negative, aerobic, non-spore forming bacteria, and the main criterion was their ability to nodulate.

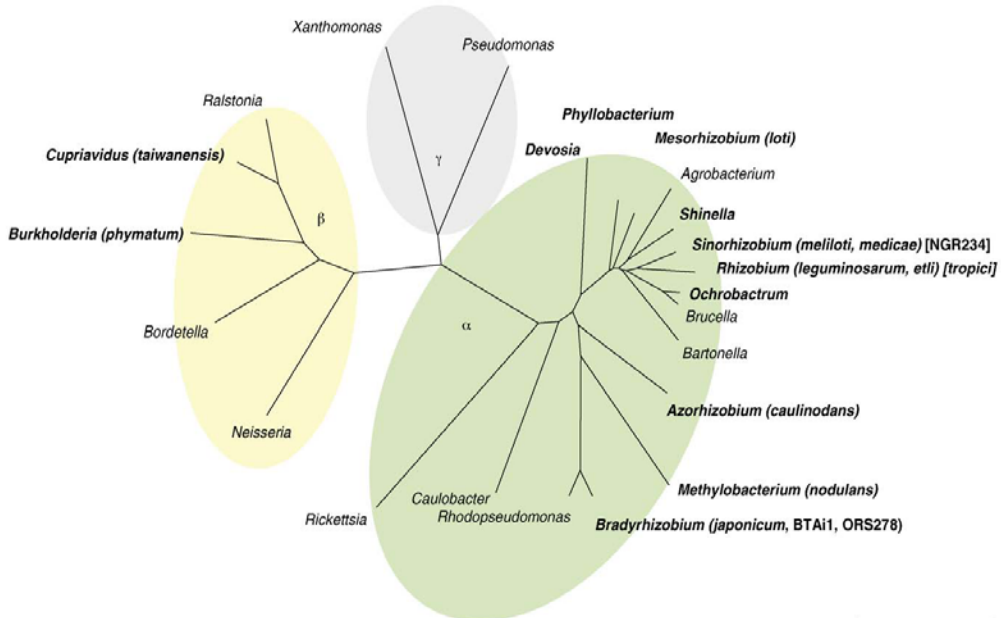


Fig. 1. Distribution of root-nodulating bacterial genera (bold font) in different classes of *Proteobacteria* in an unrooted phylogenetic tree based on 16S rDNA sequences (Masson-Boivin *et al.*, 2009). Root nodulating bacteria are distributed mainly in the classes' *α-proteobacteria* and *β-proteobacteria*. A few *γ-proteobacteria* have also been isolated from nodules. For example, bacteria belonging to *γ-proteobacteria* have been isolated from the nodules of *Arachis hypogea,* but these failed Koch's postulates under laboratory conditions (Ibañez *et al.*, 2009).

For a long time, it had been assumed that all nodulating bacteria came from the α-*proteobacteria,* but recently a number of bacteria from the *β-proteobacteria* were found to also show nodulation capacity, and recently rhizobia have been classified as α-rhizobia and β-rhizobia (Moulin *et al.*, 2002).

1.6 Taxonomy of rhizobia (α-rhizobia)

In Bergey's original manual, bacteria were included in rhizobia if they showed nodulation capacity, while bacteria with the same morphological characters that did not nodulate were excluded. Later, nodulation host range and behavior on growth media were also considered (Baldwin & Fred, 1929; Fred *et al.*, 1932) for rhizobial classification. Based on growth behavior on media, Fred *et al.* (1932) classified rhizobia as either fast growing and slow growing rhizobia (Young, 1996; and references therein).

Although the transferable nature of plasmids and plasmid-borne genes were well known to scientists (Zurkowski & Lorkiewicz, 1976; and reference therein), plasmid-borne character like nodulation capacity was dominant in rhizobial taxonomy for a long time. For example, six rhizobial species were included in the genus *Rhizobium* by Jordan and Allan (1974) in the first Bergey manual of determinative bacteriology based on their nodulation host. In the first published valid list of bacterial species (Skerman *et al.*, 1980) four rhizobial species (*Rhizobium leguminosarum*, *Rhizobium phaseoli*, *Rhizobium trifoli* and *Rhizobium meliloti*) were included, and their names were assigned based on the nodulated host. In 1975, the International Committee of Systematic Bacteriology proposed a minimum standard for describing new taxa (species). Following the code of nomenclature for bacteria, Jordan (1984) used the idea of numerical taxonomy, DNA GC content, serological response, extracellular composition, carbohydrate utilization pattern, metabolism type, bacteriophage and antibiotic susceptibility, protein composition and bacteroid type to propose the new genus *Bradyrhizobium* in the family *Rhizobiaceae* and rearrange previously described species. The previously described three species *R. leguminosarum*, *R. trifoli*, and *R. phaseoli* were combined together into the single species *R. leguminosarum*. Two other species, *R. meliloti* and *R. loti,* remained in the genus *Rhizobium* but all slow-growing rhizobia were transferred to the genus *Bradyrhizobium* (Jordan, 1984).

The international sub-committee on taxonomy of *Rhizobium* and *Agrobacterium* (http://edzna.ccg.unam.mx/rhizobial-taxonomy/node/4) is curating information about root nodulating rhizobia. According to this, 145 bacterial species that form nodules with

different legumes have been described in the genera *Agrobacterium* (Conn, 1942), *Allorhizobium* (de Lajudie *et al.*, 1998), *Azorhizobium* (Dreyfus *et al.*, 1988), *Bradyrhizobium* (Jordan, 1982; 1984), *Ensifer* (Chen *et al.*, 1988; Scholla & Elkan, 1984) and *Rhizobium* (Frank, 1889).

1.7 Revisions in rhizobial taxonomy

A valid list of bacterial species was first published in 1980 by Skerman *et al.* (1980), and from this year it became a requirement to publish or validate all newly described bacterial species in the IJSEM journal. The proposal of Jordan (1984) for the new genus *Bradyrhizobium* and the rearrangement of the previously described species were published in Bergey's manual, but not validated in the IJSEM journal. Therefore, the previously described species remained valid until 2008, when (Ramírez-Bahena *et al.*, 2008) revised the taxonomic status of *R. trifoli, R. phaseoli* and *R. leguminosarum*. They merged *R. trifoli* into *R. leguminosarum* but retained *R. phaseoli* as a separate species (Rivas *et al.*, 2009; and references therein).

1.8 Ambiguity in *Agrobacterium-Allorhizobium-Rhizobium*

All described species in the genus *Agrobacterium* are named and included in this genus on the basis of plasmid-borne characters like tumorigenesis, pathogenicity, and rhizogenicity. Plasmids and plasmid-borne genes are not stable characters but are exchangeable among closely-related and even distantly-related bacterial species or even with plants (for example *Agrobacterium tumafaciens* is very useful for transferring DNA to plants). There are several studies reporting ambiguity between *Rhizobium* and *Agrobacterium* (Graham & Parker, 1964; White, 1972; Graham, 1976; Sawada *et al.*, 1993; Young & Haukka, 1996; Willems, 2006). Nowadays, taxonomists are trying to establish a natural classification of microorganisms by describing stable characters that avoid ambiguity. Therefore, it has been proposed to merge *Agrobacterium* and *Allorhizobium* with *Rhizobium* (Young *et al.*, 2001). After several proposals for the revision of the three genera *Rhizobium*, *Allorhizobium* and *Agrobacterium*, the proposal of Young *et al.* (2001) has been accepted and *Agrobacterium* and *Allorhizobium* have now been included in the genus *Rhizobium*.

1.9 Taxonomy of β-rhizobia

Rhizobia as a group are not monophyletic but contain diverse species of phylogenetically disparate bacteria. In total, eighteen published species have been validly described as β-rhizobia so far under the genera: *Aminobacter* (Urakami *et al.*, 1992), *Burkholderia* (Yabuuchi *et al.*, 1992), *Cupriavialus* (Makkar & Casida, 1987), *Devosia* (Nakagawa *et al.*, 1996), *Herbaspirilum* (Baldani *et al.*, 1986) *Mesorhizobium* (Jarvis *et al.*, 1997), *Methylobacterium* (Patt *et al.*, 1976), *Microvigra* (Kanso & Patel, 2003), *Ochrobactrum* (Ngom *et al.*, 2004), *Phyllobacterium* (Knösel *et al*., 1984) and *Shinella* (An *et al.*, 2006).

Although β-rhizobia have only recently been described from different legumes, molecular evidence showed that as legume symbionts they have existed for 50 m illion years (Gyaneshwar *et al.*, 2011; and references therein). Different species of rhizobia from β-*proteobacteria* contain common nodulation genes (*nodABC*, *nodD*, *nifH)* that are very similar to 'traditional' rhizobial (*α-proteobacteria*) symbiotic genes. Therefore, it has been hypothesized that β-rhizobia recruited symbiotic genes (Rivas *et al.*, 2009) by horizontal gene transfer.

1.10 Molecular Phylogeny

1.10.1 Overview of molecular phylogenetics

The study of evolutionary relationships between organisms using molecular data (DNA, RNA or protein sequences) is known as molecular phylogenetics. Zuckerkandl & Pauling (1965) proposed the concept of theoretical phylogenetic reconstruction, which views macromolecules as documents of evolutionary history that may therefore help to reconstruct phylogenies. They also introduced the "molecular clock concept", i.e. the idea that random changes in nucleotide or amino acid sequences occur at more or less constant rates, so that the number of differences between two molecules with conserved function from two organisms should be approximately proportional to the evolutionary time that has passed since divergence.

A phylogenetic tree is a diagram composed of nodes and branches, with the nodes connecting any adjacent branches. Branches represent taxonomic units that could be species, populations or individuals. Relationships between taxonomic units are defined by branches in terms of descent and ancestry; this branching pattern is known as the tree topology (Graur & Li, 2000). Following a cladistic approach only monophyletic groups derived from a common ancestor, should form taxonomic units such as genera,

tribes, families, species or subspecies (Wink, 2007; and reference therein). Monophyly, paraphyly or polyphyly may be inferred from phylogenetic analyses (Henning, 1966). Monophyletic and paraphyletic groups have a single evolutionary origin. Polyphyletic groups are the result of convergent evolution, and the main characteristic used to define the group is absent in the most recent common ancestor and consist of a hodgepodge of unrelated forms (Henning, 1966).

1.10.2 Advantages of molecular data

Molecular data, especially DNA, RNA and amino acid sequences, are much more reliable for evolutionary studies than morphological and physiological data. Morphological data can be convergent and is adaptive in nature (Wink, 2007), while molecular data are strictly heritable and less prone to convergence. DNA sequence is relatively resistant to change because of various repair and proofreading systems. Description of molecular data is straightforward and defined, while morphological descriptions always contain ambiguous modifiers such as 'thin', 'reduced', 'slightly elongated', 'partially enclosed', or 'somewhat flattened'. More importantly, molecular traits evolve in a much more regular manner than morphological and physiological traits, and hence a much clearer relationship can be inferred from DNA, RNA or amino acid sequences. Nonetheless, sophisticated mathematical and statistical models have been developed to analyze molecular data (Nei & Kumar, 2000). Finally, molecular data are abundant in public databases and much more accessible than morphological data. Prokaryotes have very little morphological and phy siological variation for use in phylogenetic studies (Nei & Kumar, 2000). The aforementioned molecular characters are thus particularly important for studying the phylogenetics of microorganisms such as Bacteria and Archaea.

1.11 Phylogeny of bacteria and rhizobia

In 1930, many famous scientists (Albert Jan Kluyver, Cornelius Van Niel and Roger Yate Stanier) initially reconstructed bacterial phylogenies on the basis of morphological characters of the cells (Oren, 2010; and references therein), and during the 1940's the use of morphological and phy siological characteristics for bacterial phylogentic reconstruction had become common place, scientists started to use. Soon after, many prominent scientists [Stainer (1916 – 1989); Nile (1897 – 1985); Ernst Georg Pringsheim (1881 – 1970); Sergei Winogradsky (1856 – 1953)] already realized that morphological

and physiological properties were insufficient to construct a natural taxonomy of bacteria, but had not yet found a viable alternative (Oren, 2010; and reference therein). Because of this difficulty, this period became known as “the Dark age” of bacterial phylogeny (Woese *et al.*, 1984).

The development of protein and nucleic acid sequencing technology in 1978 revolutionized all aspects of biological research. Woese and his colleagues obtained surprising phylogenetic results using sequencing techniques, which became the ultimate starting point for the comparison of 16S and 18S rRNA genes from prokaryotes and eukaryotes (Fox *et al.*, 1977). During the 1970’s, Carl Richard Woese (1928 – 2012) proposed his famous three domain model of life based on nucleotide sequence of 16S rRNA: Archaebacteria, Eubacteria and Eukaryotes (Woese & Fox, 1977). The concept of “three domains of life” is now well accepted by the scientific community, with very few exceptions.

Due to a lack of sufficient morphological and physiological variations, and availability of sequencing facilities, microbiologists are now depending on molecular data, especially chromosomal DNA sequences, for taxonomic and systematic purposes. However, phylogenetic analyses of different genes also help to understand the evolutionary history of single genes or entire genomes of microorganisms, and hence microbiologists from the fields of systematics, taxonomy and epidemiology are using DNA sequences substantially. Nowadays, it is clear that any phylogenetic hypothesis has to be accepted or rejected on the basis of molecular sequences.

Although rhizobial characterization, classification and identification began on the basis of nodulation capacity and morphological characteristics, analyses of the 16S rRNA gene sequences later helped to eliminate plasmid-borne characters (nodulation, pathogenicity or hairy root formation) and significantly contributed to the identification of non-traditional rhizobia. Therefore, after the proposal of Woese *et al.* (1984) 16S rRNA gene sequencing became a main criterion for rhizobial taxonomy, and nodulation became secondary (Rivas *et al.*, 2009).

Bacterial taxonomists suggest different methods for identifying bacterial species.These methods include sequencing of the 16S rRNA gene, restriction fragment length polymorphism (RFLP) typing, multilocus sequence analyses of different protein-coding housekeeping genes (MLSA), whole-genome sequence analysis, fourier transformed infrared spectroscopy (FTIR) and pyrolysis mass spectrometry. However, the

sequencing of housekeeping genes, DNA profiling and the application of DNA arrays are preferred (Brenner *et al.*, 2005; and references therein). There are also some powerful PCR-based techniques like REP- and ERIC-PCR available for bacterial taxonomy, and their discriminatory power is higher than serological, RFLP and multi locus enzyme electrophoresis (MLEE) techniques (Vinuesa *et al.*, 2005). However, for the description of root- and stem-nodulating bacteria, a minimum standard has been pr oposed by (Graham *et al.*, 1991) who suggested to use a combination of traditional morphological and culture characteristics, symbiotic properties, DNA fingerprinting methods, 16S rRNA gene sequencing and DNA hybridization.

1.12 Phylogenetic inference

Evolution in different lineages does not accumulate changes at a uniform rate. In other words, different branches may evolve at different rates of change. In a realistic evolution scenario, variations in divergence rates could arise due to random changes, differences in mutation rate and selective pressure. Thus, similarity alone may not be a good indicator of the degree of relatedness. Therefore, it is important that reconstruction methods consider variation in rates of divergence between lineages (Barton *et al.*, 2007).

Bifurcation is a major assumption in phylogenetic reconstructions, but is not true in many cases and can lead misinterpretations. For example, recombination is essential in the sexual reproduction process, and combines and matches genetic variation in different lineages. Therefore, the evolutionary histories for such lineages are not perfectly vertical, and these histories cannot be r epresented correctly by bifurcating trees. Moreover, lineages that have experienced hybridization and lateral gene transfer have exchanged / transfered DNA between distinct lineages, and thus a bifurcating tree may show an incorrect interpretation. Furthermore, gene duplication, deletion, domain shuffling and gene conversion can create complex evolutionary scenarios (Barton *et al.*, 2007).

Convergent and parallel evolution can cause different lineages to become more similar to each other, at least for some features. Convergent evolution implies an independent origin of derived characters in two or more lineages and is critical for phylogenetic interpretation. Similarities between lineages that are not due to a common ancestor are known as homoplasy. A degree of similarity may also be an indication of a common ancestry and helps to infer monophyletic groups. Although shared derived character states (synapomorphy) accurately indicate common ancestry, derived character states restricted to specific lineage (autapomorphies) do not provide evidence for a relationship

with other lineages. Homologus character states mean that the character states shared by all taxa inherited from a common ancestor to its descendents without change, reflecting the real phylogeny (Futuyma, 2009). Therefore, shared derived characters states are valid evidence of monophyletic groups only are if they are uniquely derived (Barton *et al.*, 2007).

The abovementioned problems and other complexities encouraged the development of new methods for phylogenetic inference. However, none of the methods is perfect, and each has its own strengths and weaknesses. Taking all challenges and caveats into acconut, phylogenetic inference can be remarkably accurate, successful and useful (Barton *et al.*, 2007).

1.12.1 Steps in phylogenetic analyses

Four steps are necessary in phylogenetic analyses of molecular sequences. These are: the selection of molecular markers, obtaining sequences from selected molecules, a multiple sequence alignment and tree reconstruction.

1.12.2 Selection of phylogenetic markers

There are different mechanisms such as mutation, gene duplication, genome reorganization, genetic exchange through recombination, reassortment and lateral gene transfer contributing to molecular evolution in bacteria (Vandamme, 2009; and references therein).

Information from substitutions, insertions, deletions is considered for phylogenetic reconstructions. Single-copy orthologus genes are important molecular markers for tracing phylogenetic relationships. Gene duplications, loss and acquired events occur more frequently in multiple copy and non-essential genes. Hence, phylogenetic analyses of multicopy and non-essential genes may give ambiguous relationships. Phylogenetic markers with highly conserved regions also do not provide important information due to a lack of variation, and highly-variable genes are subject to repeated mutations in the same position. Hence, both are ignored for phylogenetic analyses (Chun & Hong, 2010; and references therein).

1.12.3 Sequencing of molecules

Sequencing of selected genes, DNA fragments or whole genomes is important and generally done by a Sequencer. DNA and RNA may be sequenced using targeted or random approaches. In the targeted approach, specific genes or genetic elements are selected from operational taxonomic units (OTUs) and amplified and sequenced using specific primers to obtain nucleotide sequences. In a random approach, genome sequencing methods are used to sequence random cDNA or genomic DNA regions, and genes or elements of interest are identified from the obtained sequences using computational search. For understanding genome evolution, whole genome sequencing and metagenomic sequencing have recently introduced.

1.12.4 Multiple sequence alignment and software

The quality of the sequencing data and their alignment is very important for the phylogenetic reconstruction and interpretation. Besides sequencing quality, the quality of the alignment is also crucial, because phylogenetic reconstruction and c onclusions depend heavily on correct sequence alignment (Ludwig, 2010). Direct correlations exist between alignment errors and inaccurate phylogenetic interpretation (Kumar & Filipski, 2007).

In phylogenetic inference from sequence data, each position of the sequence can be considered as independent character trait. Alignment is essential to identify homologus positions in sequences by assigning each sequence is to a separate row and arranging homologus positions in columns. An alignment therefore corresponds to a data matrix. Each column in the alignment corresponds to homologus traits, and specific residues (amino acid or nucleotide) represent the character states (Barton *et al.*, 2007).

To obtain a good alignment of homologus protein-coding gene sequences from closely related species is not difficult, as the presence of strong positive selective pressure on homologus genes hardly permits their sequences to diverge from each other. It is much more challenging to get a good alignment from distantly-related protein-coding genes and from genes that have many insertions and deletions. Different software packages are available to align sequences using different algorithms for inserting gaps. Although gaps are introduced to increase sequence similarity, they do not guarantee perfect matching. The manual and judicious insertion of gaps by a specialist is therefore important for good alignment. Several algorithms are available for sequence alignment.

Among these, Clustal-X (Chenna *et al.*, 2003) is widely used due to its ability to easily produce high quality alignments.

1.12.5 Phylogenetic tree reconstruction methods

The main objective of phylogenetic analyses is to find the evolutionary relationships of the studied organisms, which is generally represented as a phylogenetic tree. There are four different methods for reconstructing phylogenetic trees: Distance matrix, Maximum parsimony, Maximum likelihood, and Bayesian inference.

1.12.6 Distance matrix method

It is generally assumed that organisms having a recent common ancestor are more similar than organisms that have a more distant common ancestor. Hence, it is possible to deduce evolutionary relationships from the similarity of organisms. The distance method uses this general principle for tree reconstruction. It consist of two steps: calculation of genetic distances using appropriate substitution models (for example, Jukes model (Jukes & Cantor, 1969); Kimura two parameter model (Kimura, 1980); general time reversible model (Rodríguez *et al.*, 1990) and tree reconstruction using suitable clustering algorithms (UPGMA, Goodfelow, 1971); Neighbor-joining method (NJ, Saito & Nei, 1987). Several modifications have been done in the NJ method to improve the clustering performance (Gascuel, 1997; Bruno *et al.*, 2000).

The distance method for tree reconstructions has some advantages, like being easy to implement in computer programs and producing phylogenies within the shortest time (Chun & Hong, 2010). The distance method always depends on pairwise distances, which is not always favorable. The difference-based parameter D is not equivalent with the evolutionary distance (d), because D does not consider all available information from a sequence in the alignment. Moreover, most measures of D do not follow a linear scale of time or number of generations. A true evolutionary distance is more linear, and allows a clear phylogenetic inference (Barton *et al.*, 2007).

1.12.7 Maximum parsimony method

The maximum parsimony method is also widely used. It uses character-based algorithms rather than distances. According to this method, the best tree reconstruction method uses the least number of changes to explain the information in an alignment (Hall, 2008).

In this method, uninformative sites (conserved sites, sites with apomorphic changes and terminal changes) are not used for tree reconstruction.

To find the best tree, different algorithms, e.g. branch and bound algorithm (Land & Doig, 1960), are used. Several programs such as PAUP (Swofford *et al.*, 1996); PHYLIP (Felsenstein, 1985) or MEGA (Tamura *et al.*, 2011), use the maximum parsimony method for tree reconstruction. This method is however sensitive to long-branch attraction (Bergsten, 2005), which is easily detectable by other methods.

1.12.8 Maximum likelihood method

To reconstruct a tree from DNA or RNA sequences, Felsenstein (1981) developed the maximum likelihood (ML) method. This method depends on a purely statistical way of constructing phylogenetic trees. It allocates probabilities for generating a particular tree to all individual residues in a sequence alignment and hence utilizes information from all nucleotides or amino acids. Maximum likelihood analyses depend on a substitution model for the estimation of the probability of a given tree.

Swofford *et al.,* (1996) mentioned that *"Maximum likelihood methods of phylogenetic inference evaluate a hypothesis about evolutionary history in terms of the probability that a proposed model of the evolutionary process and the hypothesized history would give rise to the observed data. It is conjectured that a history with a higher probability of giving rise to the current state of affairs is a preferable hypothesis to one with a lower probability of reaching the observed state. In addition to its consistency properties, maximum likelihood is useful because it often yields estimates that have lower variance than other methods (i.e., it is frequently the estimation method least affected by sampling error). It also tends to be robust to many violations of the assumptions used in its models".* Moreover, Harrison & Langdale (2006) noted that *" the analysis starts with a specified tree derived from the input dataset (for example a NJ tree) and swaps the branches on the starting tree until the tree with the highest likelihood score (i.e. the best probability of fitting the data) is gained. This score is a function both of the tree topology and the branch lengths (number of character state changes). Likelihood analysis therefore allows an explicit examination of the assumptions made about sequence evolution*". A maximum likelihood tree can be reconstructed using different software packages like PAUP (Swofford *et al*., 1996); PHYLIP (Felsenstein, 1988) or MEGA (Tamura *et al.*, 2011).The ML is the most powerful tree reconstruction method, but it is

time consuming and is hence not used extensively in many laboratories (Chun & Hong, 2010).

1.12.9 Bayesian inference

The Bayesian inference method is interesting and powerful for phylogenetic tree reconstruction. Bayesian inference is based on the notion of posterior probabilities, which means that probabilities are estimated based on some model (prior expectations) after learning something about the data. It generates a posterior probability distribution for a parameter based on prior expectations and the likelihood of the data generated by a multiple alignment (Hall, 2008). Like the MP and ML methods, it deals with each character, nucleotide or amino acid residue.

The MrBayes software package (Ronquist & Huelsenbeck, 2003) implements the Bayesian method for phylogenetic inference. MrBayes uses the Metropolis-coupled Monte Carlo Markov Chain (MCMC) process, which can be visualized as a set of independent searches that occasionally exchange information. This program allows a search across probability valleys that would otherwise trap the search on a suboptimal hill. The final product is a set of trees that the program has repeatedly visited, which constitute the top of the hill. Although this method is also powerful, it takes long time for tree reconstruction (Nei & Kumar, 2000).

Bayesian analysis is similar to the ML method in that the user has to postulate a model of evolution (Rannala & Yang *et al*., 1996; Mau *et al*., 1999). However, ML analyses obtain the tree that maximizes the probability of observing the data that given tree while Bayesian analyses seeks the tree that maximize the probability of observing the data and model that given the tree. Most importantly, Bayesian analyses rescale likelihoods to true probabilities so that the probabilities over all trees is 1.0 (Nei & Kumar, 2000). Moreover, ML analyses try to find the single best tree, while Bayesian analyses try to find the best set of trees.

1.13 Reliability of phylogenetic tree

1.13.1 Bootstrapping

After the reconstruction of a phylogenetic tree, the reliability and quality of the tree has to be assessed to ensure that the tree accurately represents the actual relationships among OUTs (operational taxonomic units). Therefore, different methods have developed to

determine the reliability and robustness of tree topologies, such as bootstrapping and branch length estimation. The most common method is the bootstrap analysis, which can be implemented in different tree reconstruction algorithms. This method was introduced by Felsenstein (1985) and is a widely and successfully applied procedure. The bootstrapping algorithm resamples the columns of a sequence alignment and creates many new alignments by random sampling, replacing the original data set. New sets are similar to each other or the original alignment, but they are rarely absolutely identical. Multiple trees are then generated from the new sets of alignments and the statistical confidence of each branch is calculated. This value is known as the bootstrap value. Generally 200 – 2000 resamplings should be used for bootstrapping (Chun & Hong, 2010), but 1000 resamples are frequently used. A bootstrap support of 70% or higher is often considered as indicative of a reliable grouping or clustering in a phylogenetic tree. The bootstrapping method can be used in all available tree building methods (Van de Peer, 2003; Chun & Hong, 2010).

1.13.2 Tree rooting

Relationships between organisms are visualized by phylogenetic studies. In the absence of directly observable relationships between ancestors and descendents, the direction of the changes must be inferred by rooting the tree (Nei & Kumar, 2000). From an un-rooted tree, it is impossible to determine the directions of changes in traits of organisms. Although rooting is not strictly necessary, in most of the cases knowledge of the direction of changes is essential to reconstruct the evolutionary process (Graur & Li, 2000).There are two methods for molecular tree rooting: out-group rooting and duplicate gene rooting. Out-group rooting compares character states of the group of interest (in-group) with a closely-related but sufficiently distant group (out-group). The differences between in-group and out-group help to understand the direction of change in the final tree (Maddison *et al.*, 1984). Prior knowledge of the relatedness of the out-group to the in-group is required to select the out-group.

In duplicated gene rooting, sequences from one gene clade are used to root another gene clade (Simmons *et al.*, 2000). Duplicated gene rooting helps to unveil unexpected relationships among species or genes in the main clade (Brown & Doolittle, 1995; Mathews & Donoghue, 1999). However, a tree can also be rooted using evidence from DNA sequences of fossil record (Smith, 1994).

1.13.3 Model selection

Selection of DNA substitution model is crucial for statistical phylogenetic inference. DNA substitution model works like language interpreter. It translates the information in a set of sequences into phylogenetic information that can be directly interpreted by a biologist. Generally two types of error such as systematic error or bias and stochastic error or inflated variance could be arised from the selection of wrong substitution model during analysis. Very simplified model avoid / ignore naturally occurring evolutionary processes and hence produce systematic error. On the other hand use of excessively complex model in phylogenetic analysis produces stochastic effect. However, DNA substitution models are designed to reduce the effect of multiple substitutions at highly variable side in a DNA sequences. A proper model make a balance between these two challenging error. Therefore, it is important to use a good model for phylogenetic inference (Brown & Eidabaje, 2009).

1.14 Bacterial speciation and recombination

Speciation is a complex event and is the outcome of a variety of genetic and ecological processes and historical contingencies. In other words, it is the result of an interaction between forces such as recombination, migration and selective sweeps (Vinuesa *et al.*, 2005; and references therein). Recombination occurs frequently among bacteria and plays an important role in the evolution of most bacterial species, including rhizobia (Silva *et al.*, 2005; Vinuesa *et al.*, 2005; Bailly *et al.*, 2006; Maiden, 2006; den Bakker *et al.*, 2008; Tian *et al.*, 2012). New genetic material can be rapidly introduced by recombination, which can lead to faster evolution than mutation alone (Narra & Ochman, 2006; Redfield, 2001).

Bifurcation, a major assumption for most phylogenetic algorithms, is violated by the presence of recombination, and it is very hard to predict mutation rates and times to the most recent common ancestor in organisms which have histories for recombination events. Moreover, the presence of recombination events gives rise to phylogenetic incongruence in reconstructed trees (Holmes *et al.*, 1999; Didelot & Falush, 2007; den Bakker *et al.*, 2008). Alternatively, network analyses allow recombination for visualizing phylogentic ambiguity, and therefore this method also be considered to study microbial evolution (Didelot & Falush, 2007; den Bakker *et al.*, 2008; Chun & Hong, 2010). Several studies have been carried out on rhizobial population genetics, and the population-species interface could be explored using both population genetics and phylogenetics

(Silva *et al.*, 2005; Vinuesa *et al.*, 2005; Bailly *et al.*, 2006; Maiden, 2006; Tian *et al.*, 2012).

1.15 Multi locus sequence analyses

The 16S rRNA gene fulfills most of the requirements for a good phylogentic marker molecule (Ludwig & Klenk, 2005). However, they have some limitations like being highly functionally conserved, their pattern of sequence change, and branching pattern in the periphery of the tree, and presence of more than one copy in some bacterial genomes. Therefore, it has been suggested that 16S rRNA has a low phylogenetic resolving power for separating closely-related organisms (> 97% similarity) and analyses of this gene are not suitable for distinguishing prokaryotes at the species level (Ludwig, 2010). Considering the limitations of the 16S rRNA gene in bacterial taxonomy, a multi locus sequence analysis (MLSA) / multi locus sequence typing (MLST) using different protein coding genes has been preferred for studying closely-related species of bacteria, including rhizobia (Adekambi & Drancourt, 2004; Mutch & Young, 2004; Ludwig & Klenk, 2005; Naser *et al*., 2005; Konstantinidis *et al.*, 2006; Martens *et al*., 2008). Nowadays, MLSA is a widely accepted approach for the characterization of bacteria. After its introduction, it has been applied to the classification of a variety of bacteria and fungi, and has helped to develop a universally portable and reproducible method. In brief, MLST represents a universal approach for bacterial characterization (Maiden, 2006).

A small number of protein coding genes could be used as marker genes to evaluate the rRNA based phylogenetic conclusions (Ludwig & Klenk, 2005). These are translation elongation (IF2, EF-G, and EF-TU) and initiation factors, RNA polymerase sub-units, DNA gyrase, *recA*, *atpD*, and *hsp*60 (Adekambi & Drancourt, 2004; Glazunova *et al.*, 2009).

1.16 Symbiotic genes for phylogenetic analyses

The rhizobial genes that are responsible for nodulation and nitrogen fixations are commonly known as symbiotic genes. They are also known as auxiliary or accessory genes because they are not essential for rhizobial survival and reproduction. In general, all symbiotic genes, like nodulation gene clusters, have been found in close proximity to *nif* and *fix* genes. Specifically, rhizobial plasmids often carry nodulation *(nod, nol and noe)*, *nif* and *fix* genes, as well as many secretion related genes. In plasmid-less strains

or species these genes are often located in laterally transferrable genomic islands, also denominated as symbiotic islands (Black *et al.*, 2012; and references therein).

The evolutionary history of chromosomal genes and symbiotic genes can be different (Lane & Reeves, 2001) and different stories may be found from the phylogenetic analyses of nodulation genes and the 16S rRNA gene (Mergaert *et al.*, 1997; Haukka *et al.*, 1998). Among symbiotic genes, a few symbiotic genes like *nodD*, *nodC, nodA* and *nifH* genes are commonly included for the proper description of rhizobial species by phylogenetic analyses. Due to the horizontal transfer of symbiotic genes, they do not provide useful information for taxonomy, but do provide complementary information on host nodulation and nitrogen fixation (Mergaert *et al.*, 1997). The description of symbiotic genes is also useful for the proper identification of β-rhizobia and for rhizobial biogeography studies (Rivas *et al.*, 2009). For symbiovar identification of α-rhizobia/traditional rhizobia, descriptions of nodulation genes are essential. However, recently some rhizobial strains have been described that lack common nodulation genes, especially in *Bradyrhizobium* sp. (BTAi1, ORS278, Giraud *et al.*, 2007).

1.17 Nodulation mechanism

Communication between the plant host and rhizobia starts when the plant secretes flavonoids in the rhizosphere (Perret *et al.*, 2000). Different plant species secrete different flavonoids to select the appropriate rhizobia for effective symbiosis. The specificity is determined by the flavonoid-binding transcriptional regulator *nodD,* which is the master regulator of symbiosis (Fellay *et al.*, 1995).

Induction of nodulation (nod) genes by *NodD* is essential for the production and secretion of rhizobial signaling molecules known as Nod factors (NFs). Nod factors are oligosaccharides consisting of four or five beta 1,4 linked N-acetyl glucosamine residues with a fatty acid residue replacing the N-acetyl group at the non-reducing end. Nod factors may also be further modified by different molecules (Perret *et al.*, 2000; Cullimore *et al.*, 2001).The enzymes involved in the synthesis of the basic Nod factors structure are encoded by the *nodABC* genes, which are conserved in all rhizobia except strains BTAi1 and ORS278 (Giraud *et al.*, 2007). Nod factors are in turn recognized by the plant and trigger root hair curling and induce the formation of nodule primordia.

1.18 Coevolution and dispersion of rhizobia

In coevolution, two species evolve in response to one another. Coevolution has been described for plant-pathogenic and symbiotic bacteria. For pathogenic bacteria, it has been proposed that any change in a single base, combined with different recombination mechanisms like unequal crossing-over, gene conversion and transposition among R genes clusters, may lead to the random generation of new sequences in the host and pathogen developed new virulence against this new sequences. Consequently, different sets of specificities are randomly generated in the host plant and the bacterium in each center of diversity.

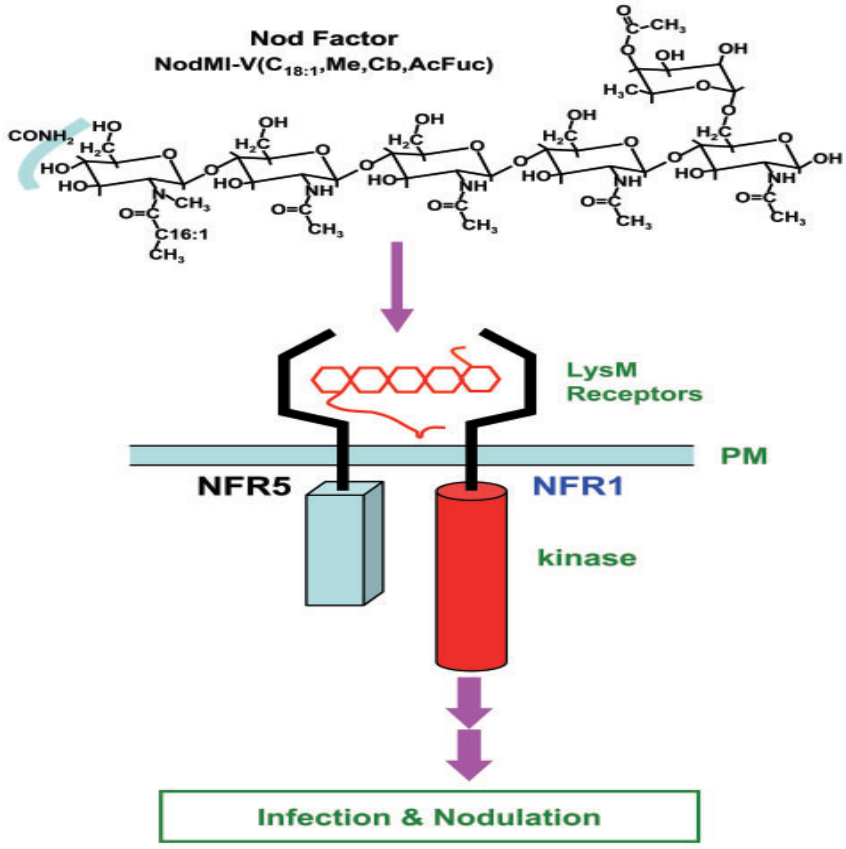


Fig. 2. Recognition of rhizobial Nod factors by plant Nod factor receptors for nodule formation. Nod factor structure determines the strict specificity between rhizobium and h ost legume, which trigger rhizobial infection process and initiation of nodule primordial in a compatible host. Five Nod factor receptors (NFR1, NFR5, LYK3, NFP, SYM10 and SYM37) have been identified in different legumes and are essential for the perception of rhizobial Nod factors for symbiosis (Kouchi *et al.*, 2010; and references therein).

Following several cycles of generation of new resistance/virulence and frequency-dependent selection, a different array of specific host-pathogen combinations appears, leading to the observed geographic differences in resistance and virulence. Host-microbe coevolution therefore occurred independently in different centres of diversification (Geffroy *et al.*, 1999; and references therein). From several experiments

(Lie *et al.*, 1987) concluded that the strong co-adaptation between legumes and indigenous rhizobial strains found in different geographical locations is a result of coevolution. Rhizobial nod factor variants induce a response in specific legume plant nod factor receptors, suggesting coevolution of legume and rhizobia (Perret *et al.*, 2000; and references therein). Moreover, Aguilar *et al.* (2004) found distinct *nodC* genotypes in rhizobia in different centres of host origin and concluded that they are distinct evolutionary lineages.

1.19 Symbiosis between rhizobia and lentil

To date, 65 valid species have been described in the genus *Rhizobium*. *R. leguminosarum* symbiovar *viciae* usually nodulates different members of the legume tribe *Vicieae* and this symbiosis may be very specific (Tian *et al.*, 2010; and references therein). In contrast, from the analysis of rhizobia from bean and pea, Santillana *et al.*, (2008) concluded that *R. leguminosarum* symbiovar *viciae* could be split into more species and proposed a taxonomic revision of this rhizobial species based on a polyphasic approach (a consensus type of taxonomy utilizing all the available data to describe consensus groups). However, most of the rhizobial strains isolated so far from legumes of the tribe *Vicieae* clearly belong to the *R. leguminosarum* species complex (*R. leguminosarum, Rhizobium fabae* and *Rhizobium pisi*) (Álvarez-Martinez *et al.*, 2009). As little work has been done on lentil symbionts, it is interesting to study the genetic diversity of lentil nodulating rhizobia from different geographical locations.

1.20 Benefit of lentil-rhizobia symbiosis

The benefit of legume cultivation has been known since ancient times. The role of root nodulating bacteria on legume growth was also observed by Hellriegel & Wilfarth (1888). Based on their own studies, Nobbe & Hiltner (1896) commercialized the effect of rhizobia on legume growth and made the first commercial rhizobial inoculum, “nitragin”, for agricultural practices (Nautiyal *et al.*, 2008; and references therein). Rhizobial inoculation has the potential to increase plant dry matter production, nitrogen yields and residual nitrogen levels of forage and grain legumes. However, the benefit of legume inoculation depends on the legume species and soil nitrogen levels (Vessey, 2004).

Similar to other legumes, lentils require infective and effective rhizobia to fix atmospheric nitrogen. Lentil showed a wide range of response to rhizobial inoculation. From FAOSTAT (2004) data and data from other sources it was concluded that the annual

fixation by lentil was about 73 kg N/ha/yr by the above-ground plant part or 110 kg N/ha/yr including the below-ground parts. The average removal of nitrogen by lentil was approximately 65 kg N/ha/yr in the harvested grain, and lentil stored 8 kg N/ha/yr in the soil for the following crop (McNeil & Materne, 2007 and references therein).

Significant responses of inoculation can be expected if lentil is grown in virgin soil. Slattery & Pearce (2002) found significant increases in nodule numbers (8-fold more), 50% more yield and 25% more dry matter yield after inoculation with acid tolerant *R. leguminosarum* strains in acidic soil (4.5 – 5), and found positive correlations between these three parameters. On the Indian subcontinent, where rhizobial populations are high, yield responses were moderate to high (Khurana & Sharma, 1995; Dhingra *et al.*, 1988; Satter, 2001; Saha *et al.*, 2008) under field conditions. Moreover, Vessey (2004) observed a significant yield response to rhizobial inoculation with lentil using commercial inoculants under field conditions. However, Moawad *et al.* (1998) investigated the competitive ability and yield response of inoculant strains in Egypt and found only 5% more yield compared to the control.

Rhizobial inoculation can significantly increase yield, seed moisture, ash, crude fiber and protein content of legumes. Regardless of whether the soil is virgin or contains endemic rhizobial populations, legume producers should inoculate their legumes for a better yield and soil nitrogen balance using low-cost rhizobial inoculum (Vessey, 2004). However, legume inoculation with effective rhizobial strains is necessary to counterbalance and reduce the use of chemical fertilizers. Therefore, it is worthwhile to manage lentil crops by inoculation to achieve maximum nitrogen fixation rates and thereby maximize total productivity and profitability (McNeil & Materne, 2007).

From the aforementioned experimental results it is clear that potential benefits from lentil inoculation with rhizobia can be achieved by using highly infective and effective strains with high survival capacity under field conditions. To find effective and infective rhizobial strains for lentil inoculation we need to isolate rhizobia from a wide range of environmental conditions. Subsequently, we need to evaluate their effect on growth under growth chamber conditions, glass house conditions and finally under field conditions to find strains suitable for use on farms.

1.21 Previous work on lentil nodulating rhizobia

Although several studies have been carried out to assess the diversity and identity of rhizobia that nodulate members of the tribe *Vicieae*, there are few reports that investigated rhizobia isolated from lentil. The studies performed on lentil rhizobia so far have mainly evaluated their symbiotic performance on plant growth and have described their biochemical characteristics and stress tolerance (salt and temperature). The diversity of rhizobia from the tribe *Vicieae*, especially from pea, faba bean and grass pea, has been studied by many scientists around the world (Laguerre *et al.*, 1996; Mutch & Young, 2004; Hou *et al.*, 2009; Tian *et al.*, 2010; Risal *et al.*, 2012; and many others), but very few studies on the diversity and taxonomy of lentil nodulating rhizobia have been carried out (Hynes & O'Connell, 1990; Moawad & Beck, 1991; Laguerre *et al.*, 1992; Geniaux & Amargr, 1993; Keatinge *et al.*, 1995). However, by analyzing rhizobia from the tribe *Vicieae* from different countries it has been concluded that *R. leguminosarum* is the main lentil-nodulating species (Tian *et al.*, 2010; and reference therein).

1.22 Research strategy

In order to increase our knowledge of the genetic diversity of rhizobial populations associated with members of the tribe *Vicieae*, specifically with lentil, and to determine the taxonomic position and identity of lentil symbionts, we isolated rhizobia from field-grown lentil nodules which were collected from different localities of Bangladesh and Germany. Lentil-nodulating rhizobia were also isolated from the nodules of lentils grown in potted soil samples under glass house conditions. These soils were collected from lentil-growing countries where this crop originated (Turkey & Syria; Lev-Yadun *et al.*, 2000; Sonnante *et al.*, 2009) and from Germany, where it introduced later. In this study, we aim to (1) explore the genetic diversity and identity of lentil nodulationg rhizobia (2) evaluate the levels of genetic diversity and the population structure of these bacteria from different geographic locations (3) to know the transmissible nature of nodulation genes. Here, concepts from prokaryotic evolutionary theories and population genetics have been used to describe our result and interpretations.

1.23 Objectives

The diversity of lentil-nodulating rhizobia has been described on the basis of plasmid profiles, RFLP and rep-PCR (repetitive element sequence-based PCR), but no sequence information is currently available (Laguerre *et al.*, 1992; Geniaux & Amargr, 1993;

Tegegn, 2006). To our knowledge, phylogenetic analyses have not yet been performed on lentil rhizobia using a MLSA approach. Therefore, it is interesting to study the genetic diversity of lentil nodulating rhizobia from different geographical locations using an MLSA approach.

The main aim of this study was to investigate the genetic biodiversity of rhizobia associated with lentil (*Lens culinaris,* Medik.) from different countries. There were several questions, *e.g.* what are the symbionts of *Lens culinaris*? Are they native to particular country/countries? Is same genotype present in different countries? Are they influenced by lentil cultivars, greges, races, sub-species, and species?

To answer these questions, lentil-nodulating rhizobia were isolated from traditional lentil-growing countries like Bangladesh, Turkey and Syria, and from Germany, where lentil is now only grown in scattered localities. A total of 134 rhizobia from lentil nodules were isolated from the four different countries. We have determined some morpho-physiological characteristics of the collected rhizobia, like biochemical properties, tolerance to acid or alkaline conditions, salt concentrations and temperature, and resistance to different antibiotics. The symbiotic properties were evaluated by conducting nodulation tests with lentil and cross-inoculation test with grass pea and pea. We have sequenced four chromosomal genes and three plasmid-borne nodulation genes from 94 isolates, and 406 sequences have deposited in the GenBank from the isolates used in this study. Nucleotide sequences have analyzed using phylogenetic and population genetic approach.

2. Materials and Methods

2.1 Sources of nodule samples and their collection localities

All nodule samples were collected from field-grown and glasshouse-grown lentil roots. All Bangladeshi isolates were collected from field grown lentil nodules and isolates from Germany were collected from both field grown and glasshouse grown lentil nodules. To grow lentil under glasshouse condition, soil samples were collected from different localities of Germany, Turkey and Syria. Nodule collection and preservation, and isolation of rhizobia were carried out following standard protocols (Somasegaran & Hoben, 1994). Nodules from lentil roots were collected from 25 localities in Bangladesh, 10 localities in Germany, two localities in Syria and one locality in Turkey (table 2 and table 3). Isolates numbers, localities and their respective geographical positions in Bangladesh are listed in the table 2. Details of the isolates from Germany, Turkey and Syria are shown in table 3.

2.2 Isolation rhizobia from nodules and their preservation

The collected nodulated roots were washed with water, dried with tissue paper and preserved on silica gel dessicant until further processing. For the isolation of bacteria from nodules, a single nodule was crushed in 50 µL of sterile water using a homogenizer. A loop-full of suspension was streaked on yeast-extract mannitol agar (YEMA) plates (Vincent, 1970) and incubated at 28°C for 3 – 5 days. Isolated single colonies were purified by repeated streaking on YEMA and CRYEMA plates (Somasegaran & Hoben, 1994). A total 134 rhizobial strains (36 from Bangladesh, 78 from Germany, 12 from Turkey and 8 from Syria) were isolated in this study. Single colonies were preserved either on agar slants at 4°C, or frozen in broth with 50% glycerol at −80 °C until further analysis. All collected bacterial isolates were preserved at the Institute of Pharmacy and Molecular Biotechnology (IPMB, Germany) and at the Bangladesh Institute of Nuclear Agriculture (BINA, Bangladesh) for further study. The names of isolates and GenBank accession numbers for the different gene sequences provided in the table 4 (Bangladeshi isolates) and table 5 (German, Turkish and Syrian isolates).

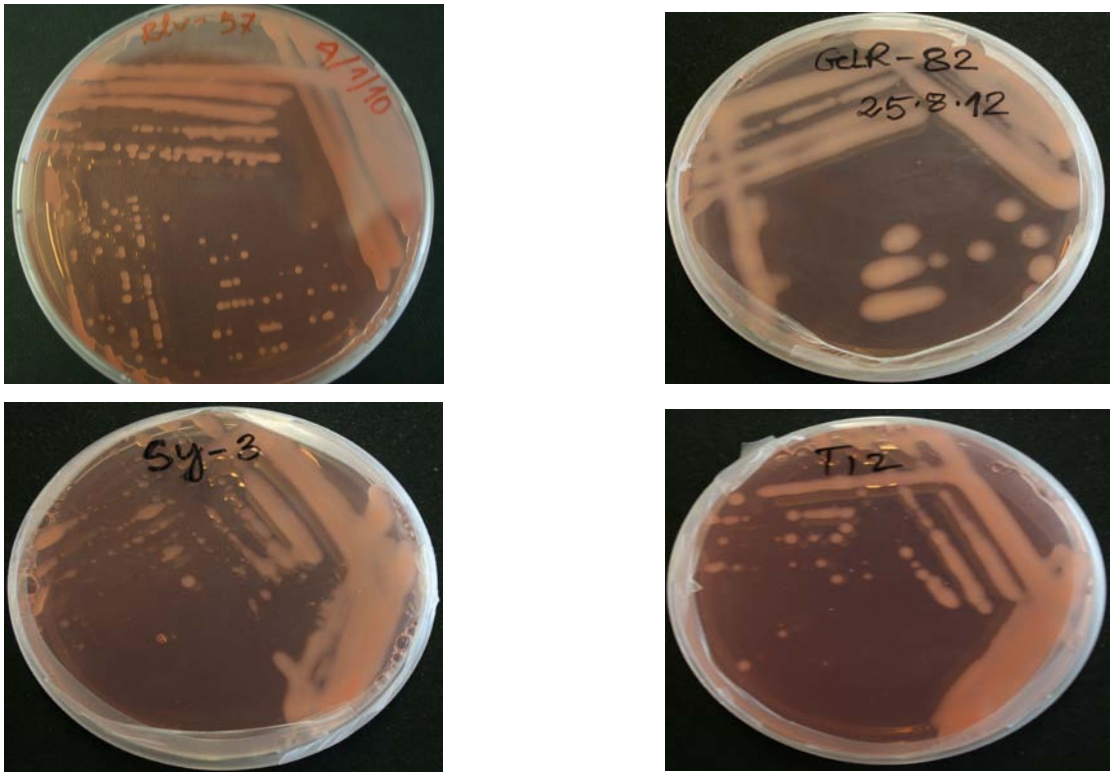


Fig. 3. Isolated single colonies of lentil-nodulating rhizobia on CRYEMA plate

2.3 Phenotypic characterization

Different phenotypic characteristics of isolated bacteria such as tolerance to high temperature, NaCl, pH, and antibiotics; size and morphology of colonies, acid production and growth on LB media were determined following standard procedures (Amarger *et al.*, 1997; Wang *et al.*, 1998; Valverde *et al.*, 2006; Bromofeld *et al.*, 2010). The tolerance to temperature, NaCl, pH and ant ibiotics resistance was determined on T Y medium. Overnight cultures (5 µL, circa 1 × 10^8 cell / mL) were used to inoculate different media. The pH was adjusted with NaOH or HCl and media were buffered with 25 mM HOMOPIPES (pH 4.5), 40 mM MES (pH 5.5), 30 mM HEPES (pH 6.8 – 8.2) (Moron *et al.*, 2005), or CHES (pH 10). Temperature tolerance was assayed by incubating inoculated plates at different temperatures (4 – 45°C). Inoculated plates containing media with different pH values, concentrations of antibiotics or NaCl, LB medium, or YEMA with BTB (bromothymol blue) were incubated at 28°C for 3 – 5 days. However, in this study, the strains *R. leguminosarum* symbiovar *viciae* 3841, *Rhizobium etli* CFN42 and *Sinorhizobium meliloti* 1021, kindly provided by Prof Dr. J.P.W. Young, University of York, England, were used as references.

Table 2. Geographical coordinates of sample collection localities in Bangladesh

Locality number	Isolates	Village/ union	Upazila/ subdivision	District/ division	Geographical coordinates
1	9, 12	Jeopara	Puthia	Rajshahi	24°24′10″ N, 88°50′0″ E
2	27, 28	Bagatipara	Bagatipara	Natore	24°21′48″ N, 88°55′42″ E
3	26	Fatehpur	Lalpur	Natore	24°12′59″ N, 88°58′30″ E
4	45, 46	Mooladooli	Ishwardi	Pabna	24°10′0″ N, 89°4′0″ E
5	39, 41	Baromile	Bheramara	Kushtia	24°3′0″ N, 88°58′0″ E
6	29, 33	Mirpur	Mirpur	Kushtia	23°58′0″ N, 89°0′0″ E
7	100	Nangal Jara	Kalaroa	Satkhira	22°53′15″ N, 89°2′30″ E
8	99	Labsha	Satkhira Sadar	Satkhira	22°44′0.12″ N, 89°4′0″ E
9	105	Nagor Ghata	Tala	Satkhira	22°46′0″ N, 89°15′0″ E
10	160, 174	Ghurnia	Dumuria	Khulna	22°51″ N, 89°25′30″ E
11	175	Mohammadpur	Khulna Sadar	Khulna	22°50″ N, 89°33′0″ E
12	129	Gangni	Mollahat	Bagerhat	22°57′30″ N, 89°42′0″ E
13	153,154	Chandra Dighalia	Gopalganj Sadar	Gopalganj	23°2′0″ N, 89°49′0″ E
14	137,139	Patgati	Tungipara	Gopalganj	22°55′0″ N, 89°53′0″ E
15	122	Noapara	Fakirhat	Bagerhat	22°47′48″ N, 89°42′30″ E
15	127	Fakirhat	Fakirhat	Bagerhat	22°47′48″ N, 89°42′30″ E
16	87, 98	Kulia	Debhata	Satkhira	22°35′0.12″ N, 88°57′0″ E
17	235	Osmanpur	Mirsharai	Chittagong	22°48′0″ N, 91°34′50″ E
18	228	Mohori project	Sonagazi	Feni	22°53′0″ N, 91°23′12″ E
19	195	Sarishadi	Feni Sadar	Feni	23°3′0″ N, 91°23′45″ E
20	57, 58	Bhairab	Bhairab	Kishoregonj	24°4′0″ N, 91°0′ E
21	62	Morichar Chor	Ishwarganj	Mymensingh	24°42′0″ N, 90°37′30″ E
22	59	Boyra	Mymensingh Sadar	Mymensingh	24°46′0″ N, 90°25′0″ E
23	288	Shankarpur	Sadar	Dinajpur	25°39′45″ N, 88°35′45″ E
24	299	Bhandara	Biral	Dinajpur	25°37′59″ N, 88°31′59″ E
25	281	Kusha Ranigonj	Pirganj	Thakurgaon	25°52′15″ N, 88°20′45″ E

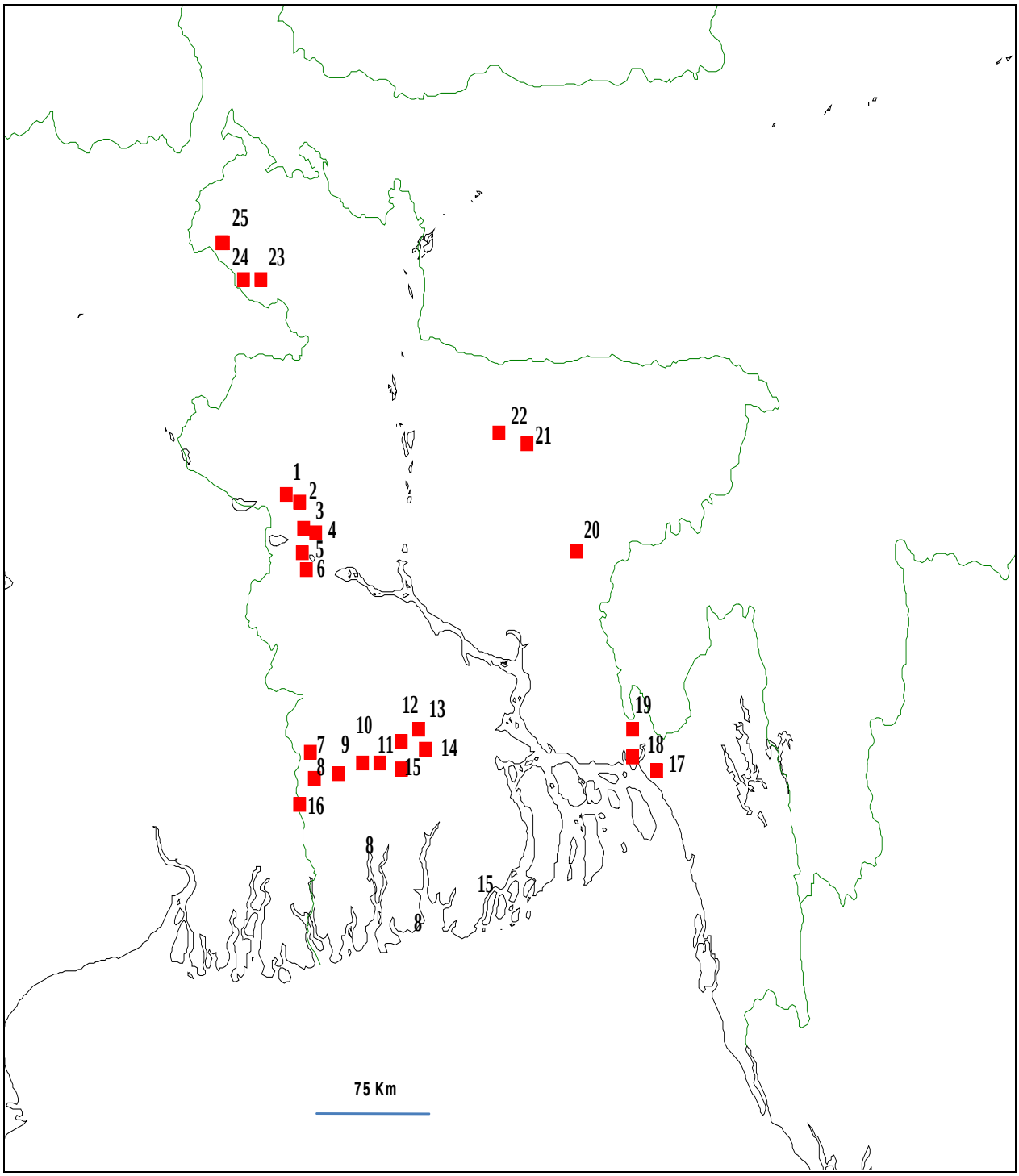


Fig. 4. Sampling localities in Bangladesh (red squares).

1: Jeopara, 2: Bagatipara, 3: Fatehpur, 4: Mooladooli, 5: Baromile, 6: Mirpur, 7: Nangal Jara, 8: Labsha, 9: Nagor Ghata, 10: Ghurnia, 11: Mohamadpur, 12: Gangni, 13: Chandra Dighalia, 14: Patgati, 15: Noapara and Fakirhat, 16: Kulia, 17: Osmanpur, 18: Mohori Project, 19: Sarishadi, 20: Bhairab, 21: Boyra, 22: Morichar Chor, 23: Shankarpur, 24: Bhandar, 25: Kusha Ranigonj.

Table 3. Name of country, isolate numbers, soil pH and rhizobial density in sample collection areas

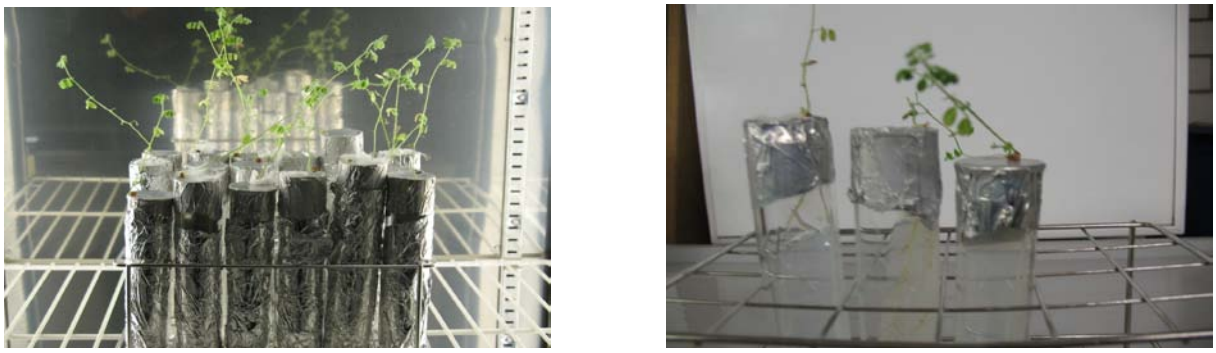
Country	Locality	Isolates	Soil pH	MPN of rhizobia (cells × g^{-1} soil)
	Bürstadt, HG	GLR1, GLR2, GLR3, GLR5	7.14	1.14×10^2
	Rittersheim, RLP	GLR6, GLR7, GLR8, GLR9, GLR10	6.92	1.14×10^2
	Plankstadt, BW	GLR11, GLR12, GLR13, GLR14	7.10	ND
Germany	Heidelberg, Wieblingen, BW	GLR16, GLR17, GLR19	7.27	ND
	Ziegelhausen, Heidebuckelweg, BW	GLR22	4.85	ND
	Babenhausen, HG	GLR23, GLR25, GLR27, GLR28, GLR29, GLR30,GLR31	6.81	5.65×10^2
	Hüfingen, BW	GLR32, GLR33, GLR34	6.78	ND
	Ostrach, BW	GLR40, GLR43, GLR45	7.25	2.18×10^3
	Heidelberg, Handschuhsheimer Feld, BW	GLR46, GLR49, GLR50	6.54	ND
	Lauterach, BW	GLR54, GLR59, GLR67,GLR69, GLR71, GLR74, GLR79	6.5-7.0	ND
Turkey	Burdur, Southwestern Turkey	TLR2, TLR3, T4, TLR5, TLR6, TLR7, TLR8, TLR9, TLR10, TLR11, TLR12, TLR14	ND	ND
Syria	Al-Hannadi, Latakia	SLR1, SLR2, SLR3, SLR4	7.58	ND
	Tishreen Suburb, Latakia	SLR5, SLR6, SLR7, SLR8	7.65	ND
Bangladesh	Rashid *et al.* (2012)	BLR9, BLR27, BLR28, BLR153, BLR175, BLR195, BLR235	ND	ND

Abbreviations, ND = not determined; BW = Baden-Württemberg, RLP = Rheinland-Pfalz, HG = Hessen, MPN = most probable numbers

2.4 Nodulation and cross-inoculation tests

All bacterial isolates were tested for nodule formation with lentil. Cross-inoculations were performed with *Lathyrus sativus* and *Pisum sativum*, two related members of the legume tribe *Vicieae*. Seeds from two lentil varieties [BINA-3 from Bangladesh and Teller Linsen from Mullers and Muhle (commercial grade) from Germany], one variety of *Lathyrus sativa* (BINA-1 from Bangladesh) and one of *P. sativum* (green pea, unshelled, commercial grade from Mullers and Muhle, Germany) were surface-sterilized using 70% ethanol (1 min) and 3% NaClO (3 – 5 min). After surface sterilization, the seeds were washed at least six times with sterile distilled water in order to completely remove the disinfectant. After imbibing (4 h in sterile water), seeds were transferred aseptically to 1% water agar plates and allowed to germinate for 2 days at room temperature in the dark.

Seedlings were later transferred to glass tubes (32 mm × 170 mm) containing Fåhraeus (1957) agar medium. Bacterial cultures (2 mL / plant) grown in YEM liquid medium (circa 1×10^8 cells / mL) were used to inoculate 5 days old seedlings (Somasegaran & Hoben, 1994). Plants were alternately irrigated with sterile de-ionized water and J enson's nitrogen-free seedling solutions. The same procedure was used for cross-inoculation assays with 10 r andomly selected isolates that were able to form nodules under laboratory conditions from Bangladeshi isolates and 30 isolates from Germany, Turkey and Syria. Plants were grown for 3 – 5 weeks in a plant growth chamber set to 25°C with 14 h light / 10 h dark cycles. Three replicates were used for each bacterial isolate in the nodulation tests. Un-inoculated plants served as negative controls.



A. B.

Fig. 5. Plant infection test sets under growth chamber conditions. A: Plants on agar based medium, B: Plant roots with nodules in agar based medium in glass tubes

2.5 Symbiotic effectivity test

For the determination of symbiotic effectiveness, an experiment was conducted under glasshouse conditions employing sterile sand. In this experiment, 17 i solates from Bangladesh, *R. leguminosarum* symbiovar *viciae* 3841, and un-inoculated controls with or without nitrogen fertilization were evaluated for lentil growth and nodul ation. Commercially available lentil seeds (Linsen from Kiepenkerl, Germany) were surface-sterilized, germinated and t ransferred to sterile sand as described above. Plant inoculation was performed as described for plant infection and cross-inoculation assays. Sterile water and nitrogen-free seedling solution were used for irrigation. In the nitrogen-fertilized control, the plants were also irrigated with KNO_3 solutions (0.5 g / L) (Somasegaran & Hoben, 1994). After four weeks, all plants were harvested and nodules were separated from the roots. The plants were dried at 70°C for 72 h, and nodules at 105°C for 48 h (Laguerre *et al.*, 2007). The data were analyzed using SPSS version 17 and the means were compared using ANOVA statistical significance tests at the 5% level and results are shown in table 10 (A – H).

A B.

Fig. 6. Plants with nodules from cross-inoculation test. A: Pea , B. Grass pea

A **B.**

Fig**.** 7**.** Symbiotic effectivity test at glasshouse, A: BINA-1, B: Linsen (commercial grade)

2.6 DNA extraction, gene amplification and sequencing

Bacterial cultures were grown at 28℃ overnight or for 24 h in tryptone-yeast (TY extract) medium (Beringer, 1974) and the DNA were isolated following the protocol of Chen and Kuo (1993). Extracted DNA was dissolved in TE buffer and the concentration was measured by UV spectrophotometry. PCR amplifications were performed with about 50 ng of template DNA. PCR conditions and primer sequences used for sequencing and DNA fingerprinting are shown in table 6.

For sequencing, PCR products were precipitated following Gonzalez and Wink (Gonzalez & Wink, 2010). Sequencing was performed using an ABI 3730 automated capillary sequencer (Applied Biosystems) with the ABI Prism Big Dye Terminator Cycle Sequencing Ready Reaction Kit version 3.1 by STARSEQ GmbH (Mainz, Germany). To confirm the observed sequences quality, both strands were sequenced from most of the isolates. In this study, a t otal of 159 sequences from 36 isolates from Bangladeshi isolates and 270 sequences from 58 strains from German, Turkish and Syrian isolates were generated and deposited in GenBank. Isolates names and accession numbers are listed in table 4 and table 5.

Table 4. Amplified genes and their accession number from Bangladeshi isolates

Isolates	16S rRNA	*atpD*	*recA*	*glnII*	*nodC*	*nodA*	*nodD*
BLR9	JN648902	JN648938	JN649028	JN648974	JN648993	JN648983	JN649014
BLR12	JN648903	JN648939	JN649029	JN648975			
BLR26	JN648904	JN648940	JN649030		JN648994		
BLR27	JN648905	JN648941	JN649031	JN648976			
BLR28	JN648906	JN648942	JN649032	JN648977		JN648984	JN649015
BLR29	JN648907	JN648943	JN649033		JN648995		
BLR33	JN648908	JN648944	JN649034		JN648996	JN648985	JN649016
BLR39	JN648909	JN648945	JN649035				
BLR41	JN648910	JN648946	JN649036		JN648997		JN649017
BLR45	JN648911	JN648947	JN649037		JN648998		JN649018
BLR46	JN648912	JN648948	JN649038				
BLR57	JN648913	JN648949	JN649039		JN648999	JN648986	JN649019
BLR58	JN648914	JN648950	JN649040				
BLR59	JN648915	JN648951	JN649041				
BLR62	JN648916	JN648952	JN649042		JN649000		
BLR87	JN648917	JN648953	JN649043				
BLR98	JN648918	JN648954	JN649044		JN649001		JN649020
BLR99	JN648919	JN648955	JN649045		JN649002	JN648987	JN649021
BLR100	JN648920	JN648956	JN649046		JN649003		
BLR105	JN648921	JN648957	JN649047		JN649004		
BLR122	JN648922	JN648958	JN649048				
BLR127	JN648923	JN648959	JN649049		JN649005		JN649022
BLR129	JN648924	JN648960	JN649050		JN649006	JN648988	
BLR137	JN648925	JN648961	JN649051				JN649023
BLR139	JN648926	JN648962	JN649052		JN649007		
BLR153	JN648927	JN648963	JN649053	JN648978	JN649008	JN648989	
BLR154	JN648928	JN648964	JN649054				
BLR160	JN648929	JN648965	JN649055		JN649009		
BLR174	JN648930	JN648966	JN649056		JN649010	JN648990	JN649024
BLR175	JN648931	JN648967	JN649057	JN648979	JN649011	JN648991	JN649025
BLR195	JN648932	JN648968	JN649058	JN648980	JN649012		
BLR228	JN648933	JN648969	JN649059			JN648992	JN649026
BLR235	JN648934	JN648970	JN649060	JN648981	JN649013		JN649027
BLR281	JN648935	JN648971					
BLR288	JN648936	JN648972					
BLR299	JN648937	JN648973		JN648982			

Table 5. Genes and their accession number from German, Turkish and Syrian isolates

Isolates	16S rRNA	*recA*	*atpD*	*gln II*	*nodC*	*nodD*
GLR1	KC679411	KC679449	KC679507	KC679562	KC679619	KC679657
GLR2	KC679412	KC679450	KC679508	KC679563	KC679620	KC679658
GLR3	KC679413	KC679451	KC679509	KC679564		
GLR5	KC679414	KC679452	KC679510	KC679565	KC679621	
GLR6	KC679415	KC679453	KC679511	KC679566		
GLR7	KC679416	KC679454	KC679512	KC679567	KC679622	KC679659
GLR8	KC679417	KC679455	KC679513	KC679568	KC679623	
GLR9	KC679418		KC679514	KC679569		
GLR10	KC679419	KC679456	KC679515	KC679570	KC679624	
GLR11	KC679420	KC679457	KC679516	KC679571	KC679625	KC679660
GLR12	KC679421	KC679458	KC679517			
GLR13	KC679422	KC679459	KC679518	KC679572	KC679626	
GLR14	KC679423	KC679460			KC679627	
GLR16	KC679424	KC679461	KC679519	KC679573	KC679628	
GLR17	KC679425	KC679462	KC679520	KC679574	KC679629	KC679661
GLR19	KC679426	KC679463	KC679521	KC679575		
GLR22	KC679427	KC679464	KC679522	KC679576	KC679630	
GLR23	KC679428	KC679465	KC679523	KC679577	KC679631	KC679662
GLR25	KC679429	KC679466	KC679524	KC679578	KC679632	
GLR27	KC679430	KC679467	KC679525	KC679579	KC679633	KC679663
GLR28	KC679431	KC679468	KC679526	KC679580		
GLR29	KC679432	KC679469	KC679527	KC679581		
GLR30	KC679433	KC679470		KC679582		
GLR31	KC679434	KC679471	KC679528	KC679583		
GLR32	KC679435	KC679472	KC679529	KC679584		
GLR33	KC679436	KC679473	KC679530	KC679585	KC679634	KC679664
GLR34	KC679437	KC679474	KC679531	KC679586	KC679635	
GLR40	KC679438	KC679475	KC679532	KC679587	KC679636	
GLR43		KC679476	KC679533	KC679588		
GLR45	KC679439	KC679477	KC679534	KC679589	KC679637	KC679665
GLR46	KC679440	KC679478	KC679535	KC679590	KC679638	KC679666
GLR49	KC679441	KC679479	KC679536	KC679591	KC679639	
GLR50	KC679442	KC679480	KC679537	KC679592	KC679640	KC679667
GLR54		KC679481	KC679538	KC679593		
GLR59		KC679482	KC679539	KC679594		
GLR67		KC679483	KC679540	KC679595		
GLR69		KC679484	KC679541	KC679596		
GLR71		KC679485	KC679542	KC679597		
GLR74		KC679486	KC679543	KC679598		
GLR79		KC679487	KC679544	KC679599		
TLR2	KC679443	KC679488	KC679545	KC679600	KC679641	KC679668
TLR3		KC679489	KC679546	KC679601	KC679642	
TLR4		KC679490				
TLR5		KC679491		KC679602	KC679643	
TLR6	KC679444	KC679492	KC679547	KC679603		
TLR7	KC679445	KC679493	KC679548	KC679604	KC679644	KC679669
TLR8				KC679605	KC679645	
TLR9		KC679494	KC679549	KC679606	KC679646	KC679670
TLR10		KC679495	KC679550	KC679607	KC679647	KC679671
TLR11		KC679496	KC679551	KC679608	KC679648	KC679672
TLR12		KC679497	KC679552	KC679609	KC679649	KC679673
TLR14		KC679498	KC679553	KC679610	KC679650	
SLR1	KC679446	KC679499	KC679554	KC679611		KC679674
SLR2		KC679500	KC679555	KC679612	KC679651	
SLR3		KC679501	KC679556	KC679613	KC679652	KC679675
SLR4	KC679447	KC679502	KC679557	KC679614	KC679653	KC679676
SLR5		KC679503	KC679558	KC679615	KC679654	
SLR6		KC679504	KC679559	KC679616		KC679677
SLR7	KC679448	KC679505	KC679560	KC679617	KC679655	KC679678
SLR8		KC679506	KC679560	KC679618	KC679656	KC679679
BLR195						KC679680

2.7 Genomic fingerprinting by ERIC-PCR

For high resolution PCR, we used the primers and PCR conditions following (De Bruijn, 1992) and (Gonzalez & Wink, 2010), respectively (table 6). The 25 µL reaction mixture contained 2.5 µL of 10X PCR buffer, 2.5 µL DMSO, 2 µL 45%-dATP nucleotide mixture,

40 pmol of each primer, 2 units of *Taq* polymerase (Pharmacia Biotech), 1 µCi (α-^{33}P)-dATP (Amersham Biosciences) and approximately 80 ng of genomic DNA. DNA fragments were separated by vertical PAGE (polyacrylamide gel electrophoresis) using a Base Acer Sequencer (Stratagene, La Jolla, San Diego, CA, USA) at 65 W for 2 h. The denaturing gels (45 × 30 × 0.025 cm) were prepared with 50 mL of Long Ranger ® solution, 35 µL tetramethylethylenediamine (TEMED) and 250 µL of ammonium persulfate (APS). After drying, the gel was exposed to X-ray film for 48 h.

Table 6. Primers used for amplification of different genes

Primer	Sequence (5´ – 3´)	Gene /sequence	PCR conditions	Reference
fD1	AGAGTTTGATCCTGGCTCAG		5 min 95°C, 30 × (1 min 95°C, 1 min 55°C,	Weisburg *et al.*, (1991)
rD1	AAGGAGGTGATCCAGC	16S rRNA	1.5 min 72°C), 10 min 72°C	
63F	ATCGAGCGGTCGTTCGGCAAGGG		5 min 95°C, 30 × (1 min 94°C, 1 min 65°C *,	Gaunt *et al.*,(2001)
504R	TTGCGCAGCGCCTGGCTCAT	**recA*	1 min 72°C), 10 min 72°C	
273F	SCTGGGSCGYATCMTGAACGT		5 min 95°C, 30 × (1 min 94°C, 1 min 65°C *,	Gaunt *et al.*,(2001)
771R	GCCGACACTTCCGAACCNGCCTG	**atpD*	1 min 72°C), 10 min 72°C	
TsglnIIf	AAGCTCGAGTACATCTGGCTCGACGG		5 min 95°C, 30 × (45 s 95°C, 30 s 58°C, 1.5 min	Stepkowski *et al.*,(2005)
TsglnIIr	SGAGCCGTTCCAGTCGGTGTCG	*glnII*	72°C), 7 min 72°C	
nodCF	AYGTHGTYGAYGACGGTTC		5 min 95°C, 30 × (1 min 95°C, 1 min 55°C,	Laguerre *et al.*, (2001)
nodCI	CGYGACAGCCANTCKCTATTG	*nodC*	1.5 min 72°C), 10 min 72°C	
NBA12	GGATSGCAATCATCTAYRGMRTGG		5 min 95°C, 30 × (1 min 95°C, 1 min 55°C,	Laguerre *et al.*, (1996)
NBF12⁻	GGATCRAAAGCATCCRCASTATGG	*nodD*	1.5 min 72°C), 10 min 72°C	
nod-A1	TGCRGTGGAARNTRNNCTGGGAAA		5 min 95°C, 30 × (1 min 95°C, 1 min 60°C,	Haukka *et al.*, (1998)
nod-A2	GGNCCGTCRTCRAAWGTCARGTA	*nodA*	1 min 72°C), 10 min 72°C	
ERIC 1R	ATGTAAGCTCCTGGGGAT		5 min 95°C, 30 × (30 s 94°C, 1 min 52°C, 8 min	De Bruijn, (1992)
ERIC 2	AAGTAAGTGACTGGGGGTGAGC	ERIC-sequence	65°C), 16 min 65°C	

* Annealing temperature for *recA* and *atpD* genes for German, Turkish and Syrian isolates was 60°C

The bands were scored as 1 (present) or 0 (absent). Similarity matrices, neighbor-joining (NJ) bootstrap consensus phylogenetic trees and principal coordinate analysis (PCoA) plots were generated using FAMD version 1.25 (Schlüter & Harris, 2006).

2.8 Processing of sequence data

Obtained sequences were aligned using Bio-Edit with careful manual curation and following the recommended precautions.

2.9 Phylogenetic Analyses

The sequences were aligned with BioEdit (Hall, 1999). Phylogenetic trees were reconstructed using the neighbor-joining (NJ) algorithm (Saito & Nei, 1987) and maximum likelihood methods (ML) in MEGA version 5 (Tamura *et al.*, 2011) using the

Kimura two-parameter model K2P (K2P, Kimura, 1980) and GTR model (Tavaré, 1986), and Bayesian Inference (BI) in MrBayes version 3.1.2 (Ronquist & Huelsenbeck, 2003). Bootstrap support for each node was evaluated with 1000 replicates. In BI analyses, two inde-pendent runs of 10,000,000 generations each were performed along with four Markov chains. Trees were sampled every 500 generations and the first 5000 samples were discarded as 'burn-in'. Phylogenetic trees were reconstructed based on single as well as combined gene data sets. Trees were rooted using *Bradyrhizobium japonicum* and *Rhizobium yanglingense* as out group except for the *nodC* (Fig. 27) and *nodD* (Fig. 28) trees. Uncorrected genetic distances (p, Nei, 1987) within and between lineages and sub-lineages were calculated using MEGA 5 (Tamura *et al.*, 2011). Tree topologies were compared using the S-H test (Shimodaira & Hasegawa, 1999).

2.10 Population genetic analyses

Parameters such as recombination events were measured with DNASP version 5.10.01 (Rozas *et al.*, 2010). The population structure was evaluated with STRUCTURE version 2.3.3 (Pritchard *et al.*, 2000; Falush *et al.*, 2003). The most likely number of clusters (K = 1 – 10) was determined under an ' admixture' model, 20,000 'burn-in' and 40,000 sampling iterations and the corresponding (highest) posterior probability value associated to the data set for STRUCTURE analyses. With an estimated K = 9, five extra long runs of 20,000 'burn-in' and 100, 000 sampling iterations were performed. Hierarchical Analysis of Molecular Variance (AMOVA, Excoffier *et al.*, 1992) was conducted in ARLEQUIN version 3.5 (Excoffier *et al.*, 2005).

2.11 Recombination and mutation analyses

Levels of recombination, mutation rates and 50% majority rule consensus trees (with and without recombination) were estimated based on the combined data set using CLONALFRAME version 1.1 (Didelot & Falush, 2007). Three independent runs were performed with a 100,000 'burn-in' and 300,000 sampling iterations. Satisfactory MCMC convergence was judged following the criterion of Gelman & Rubin (1992). We made use of three approaches in order to estimate the level of recombination in the isolates from Germany, Turkey and Syria: (1) minimal intragenic recombination events (Rm) were detected and compared with expected values of coalescence simulations based on 10,000 genealogy replications at 95% confidence level (Hudson *et al.*, 1992; Rozas *et al.*, 2010) analyzing single genes and the combined data set in DNASP version 5.10.1

(Rozas *et al.*, 2010; (2) the Shimodaira-Hasegawa (S-H) test (Shimodaira & Hasegawa, 1999) was performed to compare ML tree topologies for phylogenetic congruence as implemented in TREE-PUZZLE version 5.2 (Schmidt *et al.*, 2003); and (3) recombination rates were determined by the relative impact of recombination as compared with point mutation in the genetic diversification of the lineages (r/m proportion; (Guttman & Dykhuizen, 1994) and the relative frequency of the occurrence of recombination as compared with point mutation in the history of the lineage (ρ/θ proportion; Milkman & Bridges, 1990); these analyses were carried out in CLONALFRAME version 1.1 as described before.

2.12 Apparatus, instruments, chemicals, solutions used in this study

Table 7. List of instruments and apparatuses used for this study

Instruments	Company
Automated sequencer	ABI 3100, Applied Biosystems
Autoclave, large	Webeco, Germany
Autoclave, small	Vienna, Austria
DU 640 Photometer	Beckman ,USA
Centrifuge, 1-15K	Sigma, Germany
Electrophoresis power supply unit-E452	Fröbel , Germany
Falcon tube (25, 50 mL)	Sarstedt, Germany
Freezers (-20°C, -70°C)	AEG, Santo, Liebherr Revco
Gel casting chamber/tray	Heidelberg University, Germany
Gloves	VWR international, USA
Incubator with shaking	New Brunswick Scientific, USA
Incubators (28°C, 37°C, 65°C)	Heraeus, Germany
Laminar flow: LF1800	Fröbel labortechnik, Germany
Microcentrifuge-biofuge 13R	Heraeus, Germany
Microcentrifuge: Biofuge Fresco	Heraeus, Germany
PCR machines: Tgradient thermo cycler	Senso Quest,Biometra, Germany
pH meter: Pipetman	Gilson, France
Plant growth chamber	Rubarth, Germany
Reaction tubes (0.2, 0.5, 1.5, 2 mL)	Eppendorf, Germany
Shaker: schuettler-MT4	IKA, Germany
Sterile filter (0.22, 0.45 µm)	Sartorius, Germany
SW22 shaking water bath	Julabo, Germany
UVP,Benchtop variable transilluminator	NY, USA
UV-Photometer	WPA, Hong Kong
Vertical gel rig for PA glass	Stratagene, La Jolla, San Diego, USA
Vortex mixer, genie-2	Bender & Hobein AG, Switzerland
X-ray film	Fuji , Japan

Table 8. Chemicals, enzymes, and other materials used in this study

Chemicals, enzymes and other materials	Company / Country
Acetic acid	Merck, Germany
Agarose	HYBAID-AGS, Germany
Ammonium sulfate	Gerbu, Germany
Ammonium acetate	Merck, Germany
Antibiotics (ampicillin, kanamycin, tetracyclin, nalidixic acid, chloramphenicol, streptomycin)	Sigma-Aldrich, Serva, AppliChem, Germany
Agar	Sigma-Aldrich, Germany
Big Dye Terminator kit	Applied Biosystems
Bovine serum albumin	Sigma-Aldrich, Germany
Bromo thymol blue	Serva, Germany
Chloroform	Fluka, Switzerland
DMSO	Sigma-Aldrich, Germany
Ethanol, absolute	Merck and Becker , Germany
EDTA	Roth, Germany
Ethidium bromide	Serva, Germany
Formamide	Applied Biosystems
Glycerol	Sigma-Aldrich, Germany
Gram staining kit	Sigma-Aldrich, Germany
Isopropanol	Applichem, Germany
ß-mercaptoethanol	Merck, Germany
Mannitol	Becton, USA
Mineral oil	Sigma-Aldrich, Germany
Nucleotides	Eurofins, Germany
Phenol	Merck, Germany
Potassium nitrate	Merck, Germany
Proteinase K	Merck , Germany
Sodium dodecyl sulfate (SDS)	Applichem, Germany
Sodium acetate	Merck , Germany
Taq^{TM} DNA polymerase (red taq)	Sigma-Aldrich, Germany
Taq DNA polymerase	Sigma-Aldrich, Bioron
Taq DNA polymerase	Bioron, Germany
Tris-HCl	Roth, Germany
Ultra pure water	Merck, Germany
Yeast extract	Becton, USA

Table 9. Buffers, medium and solutions used in this study

Buffers and stock solutions	Components / Company
Agarose gel solution- 1-1.5%	Agarose in water, contains 1 µg/mL ethidium bromide
Ammonium acetate- 4M	Amonium acetate ($C_2H_3O_2NH_4$) in water
Ammonium persulfate-10%	Amonium sulphate $(NH_4)_2S_2O_8$ in water
BTB solution	Bromothymol blue (stock; 0.5 g /100 EtOH); 10 mL per litre YEMA medium (concentration: 25 ppm BTB in YEMA)
Chloroform: isoamyl alcohol	$CHCl_3/C_5H_{12}O$ in ratio 24:1
CHES (pH 10)	Sigma-Aldrich, Germany
CAPS (pH 10)	Sigma-Aldrich, Germany
Congo Red solution	Congo Red (stock:0.25 g/100 mL water): 10 mL per litre YEMA medium (concentration: 25ppm BTB in YEMA)
CRYEMA medium	Mannitol ($C_6H_8(OH)_{6)}$ -10 g, yeast extract-1.0 g, KH_2PO4-0.5 g, $MgSO_4.7H_2O$ - 0.2 g, NaCl - 0.2 g, agar-18 g, congo red-10 mL, in 1000mL water, pH-7.0
DNA size marker (100 bp)	Fermentas-Thermo Fisher Scientific, USA
DNA size marker (50 bp)	Fermentas-Thermo Fisher Scientific, USA
EDTA buffer	10% EDTA, 0.5% NaF, 0.5% thymol, 1% Tris (pH 7.5)
Ethidium bromide	0.00001% in 1% agarose
Fahraeus N-free medium	$CaCl_2$- 0.1 g, $MgSO_4.7H_2O$- 0.12 g, NaCl- 0.2 g KH_2PO4- 0.1 g, Na_2HPO_4- 0.15 g, $C_6H_{5+4y}Fe_xN_yO_7$ - 0.005 g and trace element solution-1mL in 1000 mL water.
HOMOPIPES (pH 4.5)	Sigma-Aldrich, Germany
HEPES (pH 6.8–8.2)	Sigma-Aldrich, Germany
Jensen N-free solution	$CaHPO_4$- 1.0, K_2HpO_4- 0.2g, $MgSO_4.7H_2O$- 0.2 g, NaCl- 0.2 g, $FeCl_3$- 0.1 g, trace element solution-1 mL in 1000 mL water
Lysis buffer	40 mM Tris-acetate, pH 7.8, 20 mM sodium acetate, 1 mM EDTA, 75 mM NaCl, 1% SDS
Lysogeny broth	Tryptone- 10 g, Yeast extract- 5g, NaCL- 10g per litre water
MES (pH 5.5)	Sigma-Aldrich, Germany
Nucleotide mix	2.5 mM dATP, 2.5 mM dCTP, 2.5 mM dGTP, 2.5 mM dTTP
PCR buffer (10X)	100 mM Tris-HCl, 500 mM KCl, 5% TritonX-100, 15 mM $MgCl_2$ (pH 8.5)
Phenol/chloroform/isoamyl alcohol	C_6H_5OH/ $CHCl_3/C_5H_{12}O$ in ratio 25:24:1
SDS solution- 10%	$C_{12}H_{25}NaO_4S$ in water
Sodium chloride solution- 5M	NaCl in water (saturated)
Sodium hypochloride solution-3%	NaOCl in water
TAE buffer	40 mM Tris, 1 mM EDTA, acetic acid (pH 8.0)
TBA buffer	1 M Tris, 89 mM boric acid, 10 mM EDTA, pH 8.6
TE buffer	10 mM Tris-HCl, 1 mM EDTA (pH 8.0)
Trace element solution	Bo- 0.05%,Mn- 0.05%, Zn- 0.005%,Cu- 0.002%,pH- 6.5-7.0
Tryptone agar medium	Tryptone- 5 g, yeast extract- 3.0 g, $CaCl_2$- 0.66 g, agar- 18 g, pH-7.0 in 1000 mL water
λ-PSTI size standard	Fermentas-Thermo Fisher Scientific, USA
Yeast extract mannitol agar medium	Mannitol ($C_6H_8(OH)_6$)-10 g, yeast extract-1.0 g, K_2HPO_4-0.5 g, $MgSO_4,7H_2O$- 0.2 g, NaCl- 0.2 g, agar- 18 g, in 1000 mL water, pH-7.0

Research projects

3.1 Project 1: Genetic diversity of rhizobia nodulating lentil (*Lens culinaris*) in Bangladesh

3.1.1 Abstract

In order to determine the bacterial diversity and the identity of rhizobia nodulating lentil in Bangladesh, we performed a phylogenetic analysis of housekeeping genes (16S rRNA, *recA, atpD* and *glnII*) and nodulation genes (*nodC, nodD* and *nodA*) of 36 bacterial isolates from 25 localities across the country. Maximum likelihood (ML) and Bayesian analyses based on 16S rRNA sequences showed that most of the isolates (30 out of 36) were related to *Rhizobium etli* and *Rhizobium leguminosarum*. Only these thirty isolates were able to re-nodulate lentil under laboratory conditions. The protein-coding housekeeping genes of the lentil nodulating isolates showed 89.1 – 94.8% genetic similarity to the corresponding genes of *Rhizobium etli* and *Rhizobium leguminosarum*. The same analyses showed that they split into three distinct phylogenetic clades. The distinctness of these clades from closely related species was also supported by high resolution ERIC-PCR fingerprinting and phenotypic characteristics such as temperature tolerance, growth on acid-alkaline media (pH 5.5 – 10.0) and antibiotic sensitivity. Our phylogenetic analyses based on three nodulation genes (*nodA, nodC* and *nodD*) and cross-inoculation assays confirmed that the nodulation genes are related to those of *Rhizobium leguminosarum* symbiovar *viciae*, but clustered in a distinct group supported by high bootstrap values. Thus, our multi-locus phylogenetic analysis, DNA fingerprinting and phenotypic characterizations suggest that at least three different clades are responsible for lentil nodulation in Bangladesh. These clades differ from the *Rhizobium etli–Rhizobium leguminosarum* group and may correspond to novel species in the genus *Rhizobium*.

Key words: Rhizobium; *Lens culinaris*; Nodulation; Multi locus analysis; Fingerprinting; Phylogeny

3.1.2 Introduction

Rhizobia are a group of bacteria that have the capacity to form nodules on legume roots (and occasionally on stems) and can fix atmospheric nitrogen to partially or fully meet the nitrogen requirements of the plant. An effective symbiotic relationship between the bacteria and the plant hosts is crucial for the legume to achieve maximum growth efficiency. Lentil (*Lens culinaris*) is an important and popular legume in many countries for human and animal nutrition as well as for soil fertility management. Lentil can meet its full or partial nitrogen requirement for growth and development from its symbiotic partner. The lentil is indigenous to the Near East and Central Asia, and its history in agriculture is probably as old as that of agriculture itself. The putative progenitor of modern cultivated lentils has been distributed from this region to the other continents (Shandu & Singh, 2007; and references therein). It is believed that the cultivated lentil originated in the Turkey-Cyprus region, and that the centre for diversification is South Asia. The lentil was introduced to India by 2,500 years BC and it is commonly known as “Measure” in all Indian states (Nene, 2006; Sonnante *et al*., 2009) including Bangladesh and Pakistan.

Bangladesh also possesses a long history of lentil cultivation. To our knowledge, there is no historical record of the introduction of the lentil to Bangladesh, but it has been cultivated in this region for a long time. Bangladesh is a small South Asian country surrounded by various Indian states. These surrounding states share cultural sim-ilarities with Bangladesh, including linguistic similarities and an agricultural history. Linguistic comparisons for example show that the Hindi word ‘Masur’ is found to have the same meaning (‘lentil’) in India, Bangladesh and Pakistan. These three countries make up the Indian subcontinent, and we can assume an early beginning for the cultivation of lentils in Bangladesh from around 2,500 years BC.

R. leguminosarum symbiovar *viciae* usually nodulates legumes of the tribe *Vicieae* (Laguerre *et al.*, 2003; Santillana *et al.*, 2008; Hou *et al.*, 2009), and this symbiosis may be very specific (Tian *et al.*, 2010). From the analysis of several isolates, Santillana *et al.* (2008) concluded that the taxonomic status of *R. leguminosarum* is not clear, and proposed a taxonomic revision of this rhizobial species based on a polyphasic approach (a consensus type of taxonomy utilizing all the available data to describe consensus groups). However, most of the rhizobial strains isolated so far from legumes of the tribe *Vicieae* belong to the *R. leguminosarum* species complex (*R. leguminosarum*, *R. fabae* and *R. pisi*) (Álvarez-Martinez *et al.*, 2009). It has also been shown that certain rhizobial species can only nodulate a single species of legume and that certain legumes can only

be nodulated by a single rhizobial species. There are currently six recognized species within the genus *Rhizobium* that form nodules with the common bean (Valverde *et al.*, 2006; and references therein).

Nodulation and cross-inoculation assays are necessary to determine the host range of rhizobial species, and nucleotide sequences from nodulation genes may be used to provide complementary information. Gene transfer and rearrangement of symbiotic plasmids can occur under natural conditions, depending on the donor and recipient strains (Geniaux & Amargr, 1993; Zhang *et al.*, 2001) and therefore different strains of the same rhizobial species may carry similar or different nod genes. This incongruence is generally explained by lateral gene transfer in rhizobia (Han *et al.*, 2010; and references therein). To take account of differences among strains within the same species, different symbiovars have been described for same bacterial species. Three symbiovar (*viciae*, *phaseoli*, *trifolii*) have been described in *R. leguminosarum*, and these biovars were later also described in other species of rhizobia. The same symbiovar can also occur in different species of rhizobia (Perret *et al.*, 2000). Moreover, based on the description of pathovars in pathogenic bacteria, symbiotic variants or symsymbiovars have recently been proposed for describing the adaptive behavior of rhizobia with regard to their legume host, and different symsymbiovars should be distinguished by host ranges as well as by gene sequences (Rogel *et al.*, 2011; and references therein).

In prokaryotic identification and systematics, the analysis of genes coding for the SSU rRNA is one of the most widely used classification techniques. However, for the description of new species or higher taxonomic levels of bacteria, phylogenetic analyses based on 16S rRNA sequences should be integrated into a polyphasic approach like multilocus sequence analysis (MLSA) with phenotypic characterization and DNA fingerprinting (Mutch & Young, 2004; Ludwig & Klenk, 2005; Konstantinidis *et al.*, 2006). DNA fingerprints can be useful in determining the stability of dominant members of a community in large sampling projects (Hamady & Knight, 2009).

Although several studies have been carried out to assess the diversity and identity of rhizobia that nodulate members of the tribe *Vicieae*, there are few reports investigating rhizobia isolated from lentil. The studies performed on lentil rhizobia so far have mainly evaluated their symbiotic performance on plant growth and have described their biochemical characteristics and stress (salt and temperature) tolerance (Tegegn, 2006). The diversity of lentil-nodulating rhizobia has also been described on the basis of plasmid profiles, RFLP and rep-PCR (repetitive element sequence-based PCR), but no

sequence information is currently available (Laguerre *et al.*, 1992; Geniaux & Amargr, 1993; Tegegn, 2006). To our knowledge no phylogenetic analyses have yet been performed on lentil rhizobia using a MLSA approach.

In order to increase our knowledge of the genetic diversity of rhizobial populations associated with members of the *Vicieae*, specifically with lentil, and to determine the taxonomic position and identity of lentil symbionts, we collected field-grown lentil nodules from different localities in Bangladesh and performed polyphasic analyses on the isolated bacteria.

3.1.3 Materials and methods

Nodule collection, isolation of bacteria

Detail descriptions of the methods are available in chapter 2 (materials and methods).

Plant infection and nodule effectiveness assays

Methods are available in chapter 2.

Phenotypic characterization of isolates

Methods used for phenotypic characterization are available in chapter 2.

DNA isolation, PCR amplifications, gene sequencing and ERIC fingerprinting

Detail descriptions of the methods are available in chapter 2.

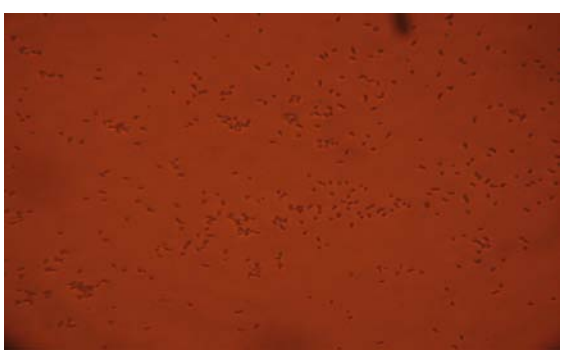
Phylogenetic analyses

Detail descriptions for phylogenetic analyses are available in chapter 2.

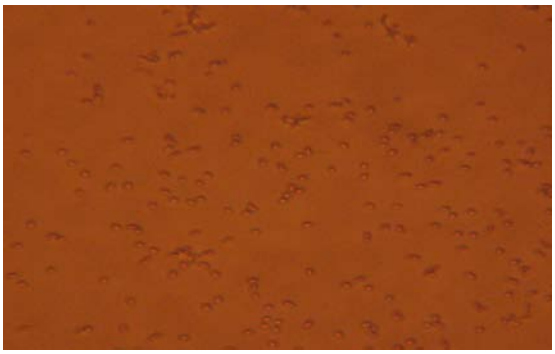
3.1.4 Results

Isolation and purification of rhizobia

Single colony from isolated rhizobia were purified by repeated streaking on CRYEMA medium as described earlier and, colony morphology and cell shape have shown in the Fig. 3 and Fig. 8 respectively.



A (40X)



B (100X)

Fig. 8. Shape of lentil-nodulating rhizobia

Nodulation and cross-inoculation

The nodulation efficiency of all isolates was tested on lentil; out of 36 isolates, 30 were able to induce the formation of nodules within 2 – 3 weeks after inoculation (Fig. 5). In cross-inoculation tests, all 10 isolates out of a randomly selected set (BLR9, BLR28, BLR45, BLR58, BLR100, BLR105, BLR154, BLR175, BLR195, and BLR235) were able to form nodules with both *L. sativus* and *P. sativum* (Fig. 6).

Symbiotic effectiveness assays

Symbiotic effectivity test were performed using sterile sand under glasshouse conditions (Fig. 7) using selected Bangladeshi isolates. Under glasshouse conditions all tested isolates [(BLR9, BLR26, BLR28, BLR41, BLR58, BLR87, BLR98, BLR99, BLR100, BLR105, BLR122, BLR129, BLR137, BLR139, BLR175, BLR228, BLR235, split into three clades (I – III), according to phylogenetic analyses, see below)], *R. leguminosarum* symbiovar *viciae* 3841 and the un-inoculated nitrogen-fertilized control treatment showed a significant effect (ANOVA, $P < 0.05$) on lentil dry matter weight production compared to the un-inoculated, un-fertilized control (table 10B). Moreover, significant differences in nodulation and plant growth were observed (ANOVA, $P < 0.05$) when comparing only the three clades (table 10D). When comparing all BLR isolates as one group with *R. leguminosarum* symbiovar *viciae* 3841, significant differences (ANOVA, $P < 0.05$) were

observed in lentil nodulation, but we did not find significant differences for plant dry matter weight production (ANOVA, P < 0.071) (table 10F).

Table 10 (A – H). Symbiotic effectivity of lentil rhizobia on growth of lentil

A. Effect of Bangladeshi isolates, *R. leguminosarum (*Rlv3841) and nitrogen on lentil growth (descriptive statistics)

Treatments	Number of observations	Plant weight (mg)			Nodule weight (mg)		
		mean	std. deviation	std. error	mean	std. deviation	mean
Clade- I	55	473.96	112.45	15.16	24.49	7.92	1.06
Clade-II	20	539.50	98.45	22.01	26.68	8.92	1.99
Clade-III	10	643.90	47.02	14.87	44.86	6.35	2.00
Rlv	5	413.20	17.76	7.94	16.06	.58	0.26
nitrogen	5	615.00	9.24	4.13	0.00	0.00	0.00
control	5	365.80	37.53	16.78	0.00	0.00	0.00
Total	100	502.67	117.33	11.73	24.09	12.65	1.26

B. Effect of Bangladeshi isolates, *R. leguminosarum (*Rlv3841) and nitrogen on lentil growth (ANOVA)

Source of variation	Plant weight (mg)			Nodule weight (mg)			df	sig.
	sum of square	mean square	F stat.	sum of square	mean square	F stat.		
Between group	468692.68	93738.53	9.85	10583.21	2116.64	37.72	5	0.00
Within group	894185.42	9512.61		5273.77	56.10		94	
Total	1362878.11			15856.98			99	

C. Effect of Bangladeshi isolates on lentil growth (clades-wise, descriptive statistics)

Treatments	Number of observations	Plant weight			Nodule weight (mg)		
		mean	std. deviation	std. error	mean	std. deviation	std. error
Clade- I	55	473.96	112.45	15.16	24.49	7.92	1.06
Clade-II	20	539.50	98.45	22.01	26.68	8.92	1.99
Clade-III	10	643.90	47.02	14.87	44.86	6.35	2.00
Total	85	509.38	117.26	12.71	27.40	10.23	1.11

Table 10 (A – H). Symbiotic effectivity of lentil rhizobia on growth of lentil (cont.)

D. Effect of Bangladeshi isolates on lentil growth (clades-wise, ANOVA)

Source of variation	Plant weight (mg)			Nodule weight (mg)			df	sig.
	sum of square	mean square	F stat.	sum of square	mean square	F stat.		
Between group	268088.12	134044.06	12.39	3523.61	1761.80	27.40	2	0.00
Within group	886945.82	10816.41		5272.42	64.29		82	
Total	1155033.95			8796.03			84	

E. Effect of Bangladeshi isolates and *R. leguminosarum* (descriptive statistics)

Treatments	Number of observations	Plant weight (mg)			Nodule weight (mg)		
		mean	Std. deviation	Std. error	mean	std. deviation	std. error
Bangladeshi isolates	85	509.38	117.26	12.71	27.40	10.23	1.11
Rlv3841	5	413.20	17.76	7.94	16.06	0.58	0.26
Total	90	504.03	116.11	12.24	26.77	10.28	1.08

F. Effect of Bangladeshi isolates and *R. leguminosarum* (ANOVA)

Source of variation	Plant weight (mg)			Nodule weight (mg)			df	Sig.	
	sum of square	mean square	F stat.	sum of square	mean square	F stat.		plant	nodule
Between group	43680.14	43680.14	3.32	607.76	607.76	6.07	1	0.07	0.01
Within group	1156296.75	13139.73		8797.39	99.97		88		
Total	1199976.90			9405.15					

Table 10 (A – H). Symbiotic effectivity of lentil rhizobia on growth of lentil (cont.)

G. Effect of rhizobial inoculation over control on lentil growth (descriptive statistics)

Isolates/ nitrogen	Replication	Plant weight (mg)			Nodule weight (mg)		
		mean	std. deviation	std. error	mean	std. deviation	std. error
BLR-9	5	377.20	38.38	17.16	26.14	2.96	1.32
BLR-26	5	425.60	25.96	11.61	41.26	4.82	2.15
BLR-41	5	415.40	34.23	15.30	14.28	1.29	0.58
BLR-58	5	480.60	54.56	24.40	22.50	3.04	1.36
BLR-87	5	602.60	24.31	10.87	27.80	3.11	1.39
BLR-98	5	389.00	11.81	5.28	33.52	3.18	1.42
BLR-100	5	585.00	47.89	21.41	18.32	2.78	1.24
BLR-105	5	318.00	21.09	9.43	17.12	2.94	1.31
BLR-122	5	380.80	35.78	16.00	21.98	2.90	1.30
BLR-137	5	624.40	25.62	11.46	22.52	2.46	1.10
BLR-139	5	615.00	22.77	10.18	23.98	4.52	2.02
BLR-62	5	521.20	47.79	21.37	18.32	4.36	1.95
BLR-99	5	545.80	57.84	25.86	24.42	4.31	1.93
BLR-129	5	421.60	25.18	11.26	39.74	2.41	1.08
BLR-175	5	669.40	28.99	12.96	24.26	4.57	2.04
BLR-228	5	645.00	645.00	15.70	40.56	4.50	2.01
BLR-235	5	642.80	642.80	27.34	49.16	4.91	2.20
Rlv3841	5	413.20	413.20	7.94	16.06	0.58	0.26
Nitrogen	5	615.00	615.00	4.13	0.00	0.00	0.00
control	5	365.80	365.80	16.78	0.00	0.00	0.00
Total	100	502.67	502.60	11.73	24.09	12.65	1.26

H. Effect of rhizobial inoculation over control treatment on lentil growth (ANOVA)

Source of variance	Plant weight (mg)			Nodule weight (mg)			df	sig.
	sum of square	mean square	F stat.	sum of square	mean square	F stat.		
between group	1258438.11	66233.58	50.73	14962.61	787.50	70.44	19	0.00
within group	104440.00	1305.50		894.37	11.18		80	
Total	1362878.11			15856.98			99	

16S rRNA gene

From 36 isolates, 1040 – 1400 base pairs (bp) were obtained from the 16S rRNA gene. The maximum likelihood (ML) phylogenetic tree based on 16S rRNA reveals that most of the isolates (approximately 83%) cluster together with, among others, *R. etli* CFN42, *R. leguminosarum*, *R. fabae* and *R. pisi* (Fig. 9). A BLAST search with these sequences shows a high similarity (99 – 100%) to *R. etli* (30 isolates). The genetic similarity (p-distance) inferred from the 16S rRNA sequences of lentil-nodulating bacteria was 99.8% to *R. etli* CFN 42 and 98.9% to *R. leguminosarum* USDA 2370 (table 11). The non-nodulating isolates were closely related to *Ensifer sp*., *R. huautlense*, *R. giardinii*, *R. tumefaciens* and *R. radiobacter*.

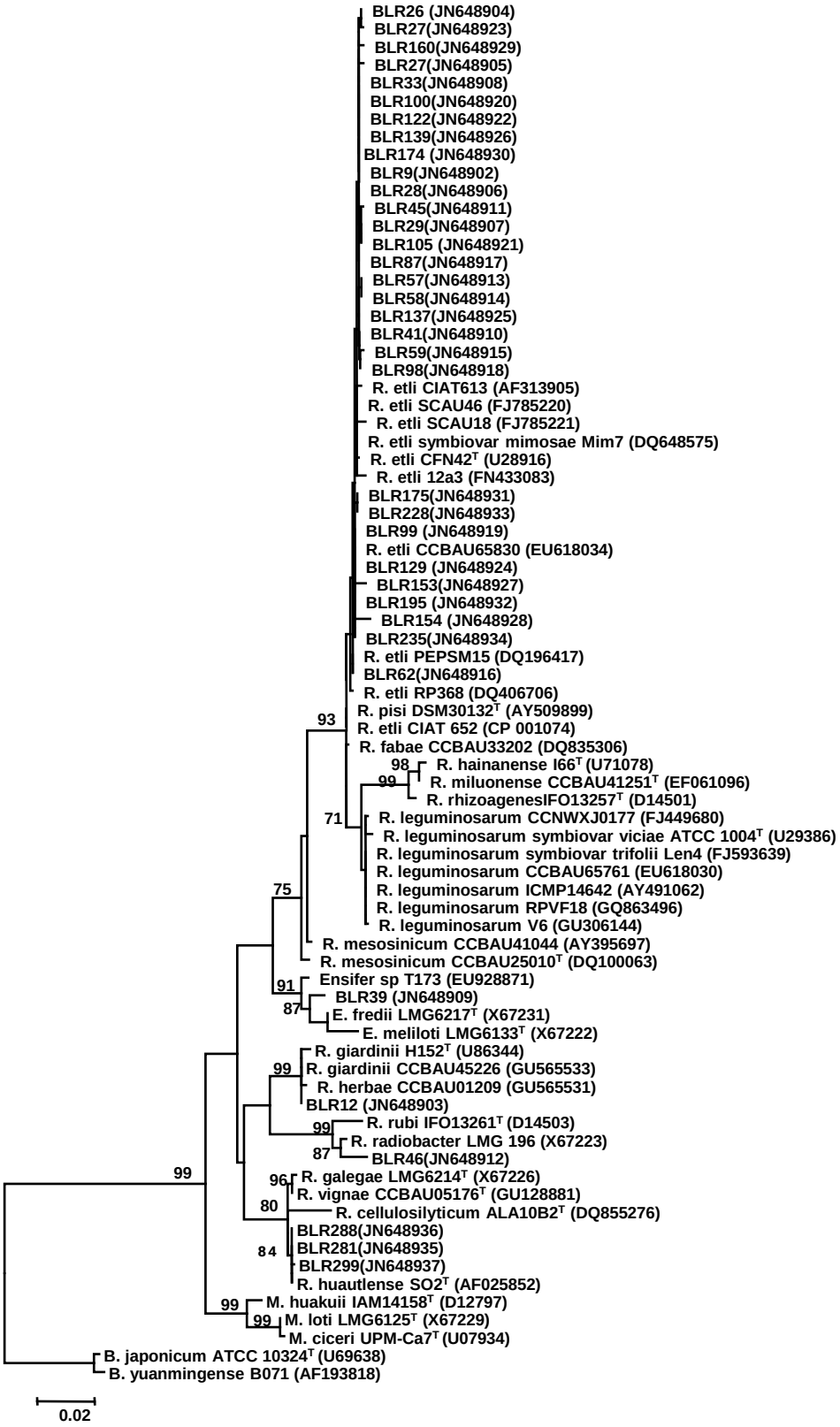


Fig. 9. ML tree based on 16S rRNA gene partial sequences. Bootstrap values indicated when ≥ 70% (1000 replicates). Abbreviations used: BLR: Bangladeshi lentil rhizobia, R: *Rhizobium,* E: *Ensifer,* B: *Bradyrhizobium,* M: *Mesorhizobium.*

Protein-coding housekeeping genes

We were able to amplify approximately 500 bp of the partial *recA* gene from 33 isolates with the primers described previously (Gaunt *et al.*, 2001) but were unable to amplify a fragment from the three remaining isolates (BLR281, BLR288 and BLR299). Considering the *recA* gene, lentil rhizobia showed 91.7 and 92.1% similarity to *R. leguminosarum* and *R. etli*, respectively. The partial *atpD* gene (approximately 550 bp) showed 89.1% similarity to *R. leguminosarum* and 94.4% to *R. etli* CFN42. The partial *glnII* gene sequences (approximately 750 bp) showed 92.8% similarity to *R. leguminosarum* USDA2370 and 94.8% to *R. etli* CFN42 (table 10). In the ML trees, based on the partial *recA, atpD* and *glnII* gene sequences, lentil-nodulating Isolates form three separate clades (I, II and III, see Figs. 10 – 12). These isolates form monophyletic groups that differ from *R. etli*, *R. leguminosarum*, *R. fabae* and *R. pisi*. The ML trees based on concatenated sequences (16S rRNA + *recA* + *atpD* and 16S rRNA + *recA* + *atpD* + *glnII*) revealed congruent topologies (Fig. 16 and Fig. 17).

Table 11. Average genetic similarity among clades and to *R etli, R. leguminosarum*

Isolates / clades	16S	*atpD*	*recA*	*glnII*
***R. etli* / nodulating isolates**	99.8	94.4	92.1	94.8
***R. leguminosarum* / nodulating isolates**	98.9	89.1	91.7	92.8
***R. etli* / Clade I**	99.8	94.5	91.4	94.8
***R. etli* / Clade II**	99.7	94.4	96.4	94.7
***R. etli* / Clade III**	99.8	93.8	88.8	94.8
***R. leguminosarum* / Clade I**	98.9	89.3	92.1	92.0
***R. leguminosarum* / Clade II**	98.8	88.6	90.4	93.1
***R. leguminosarum* / Clade III**	99.0	88.7	91.1	93.7
Clade I / Clade II	99.7	93.8	90.8	95.1
Clade I / Clade III	99.8	94.0	91.7	96.2
Clade II / Clade III	99.8	93.6	89.1	94.3

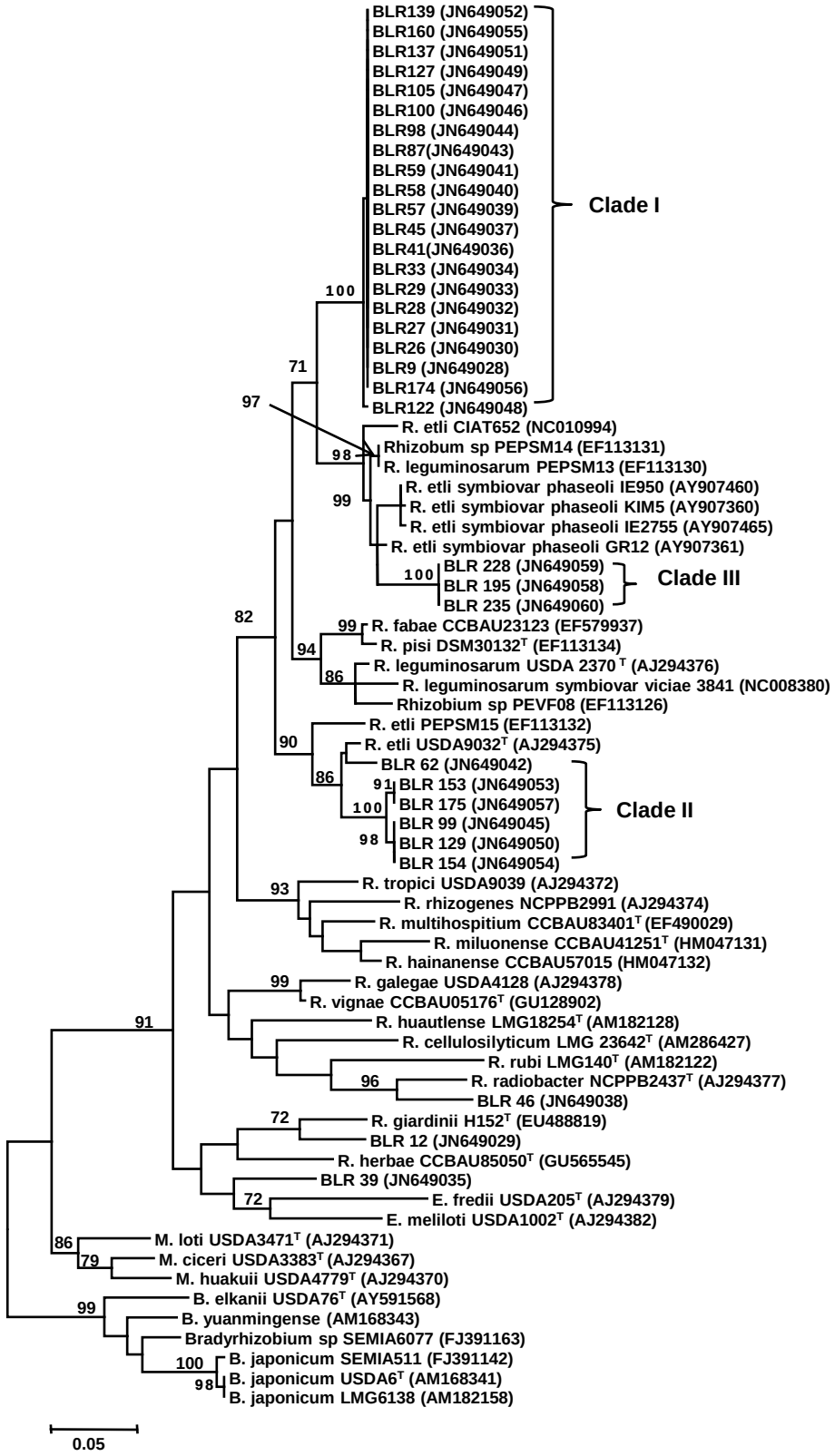


Fig. 10. ML tree based on *recA* gene partial sequences. Bootstrap values indicated when ≥ 70% (1000 replicates). Abbreviations: BLR: Bangladeshi lentil rhizobia, R: *Rhizobium,* E: *Ensifer,* B: *Bradyrhizobium,* M: *Mesorhizobium.*

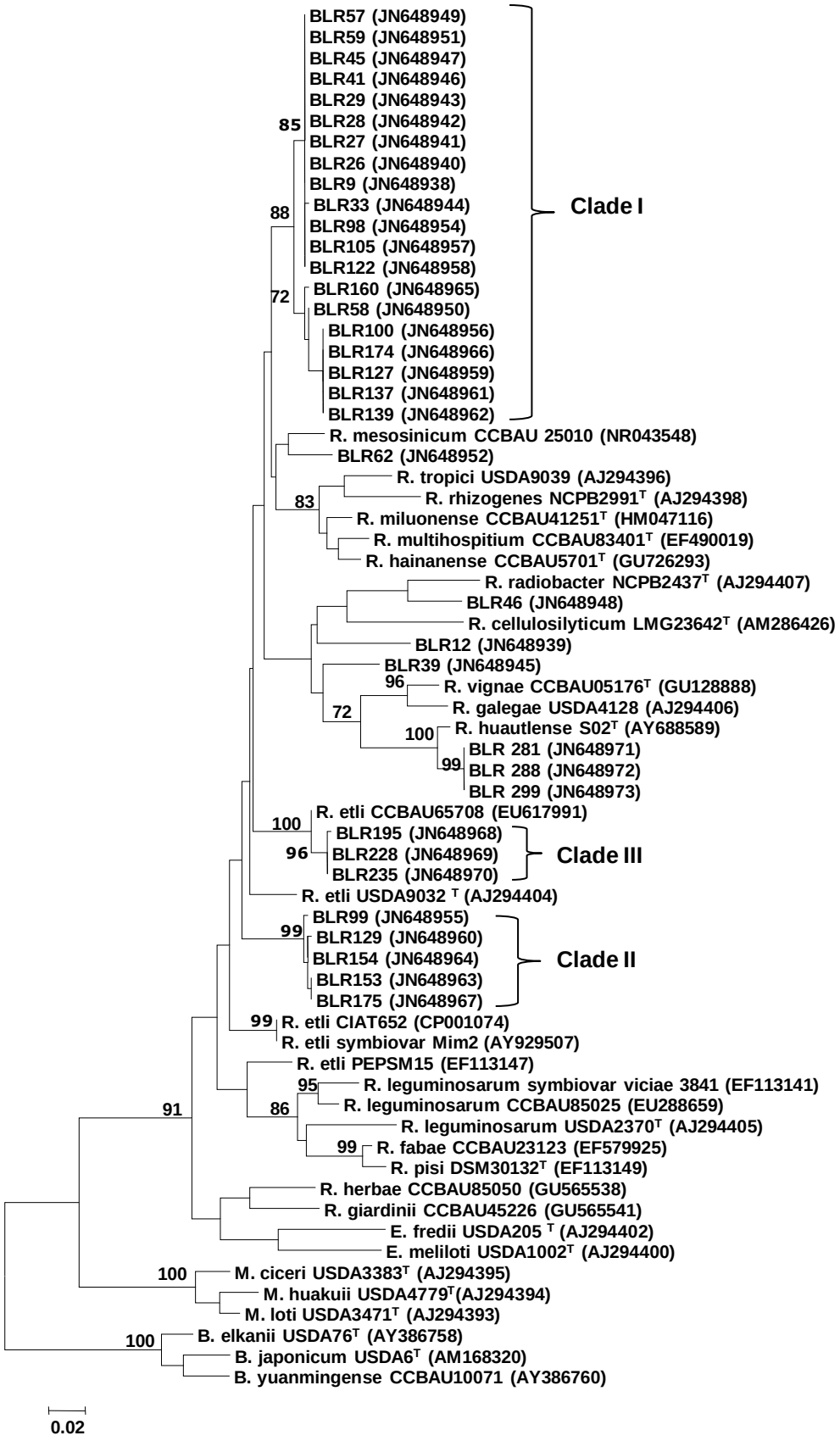


Fig. 11. ML tree based on *atpD* gene partial sequences. Bootstrap values indicated when ≥ 70% (1000 replicates). Abbreviations: BLR: Bangladeshi lentil rhizobia, R: *Rhizobium,* E: *Ensifer,* B: *Bradyrhizobium,* M: *Mesorhizobium.*

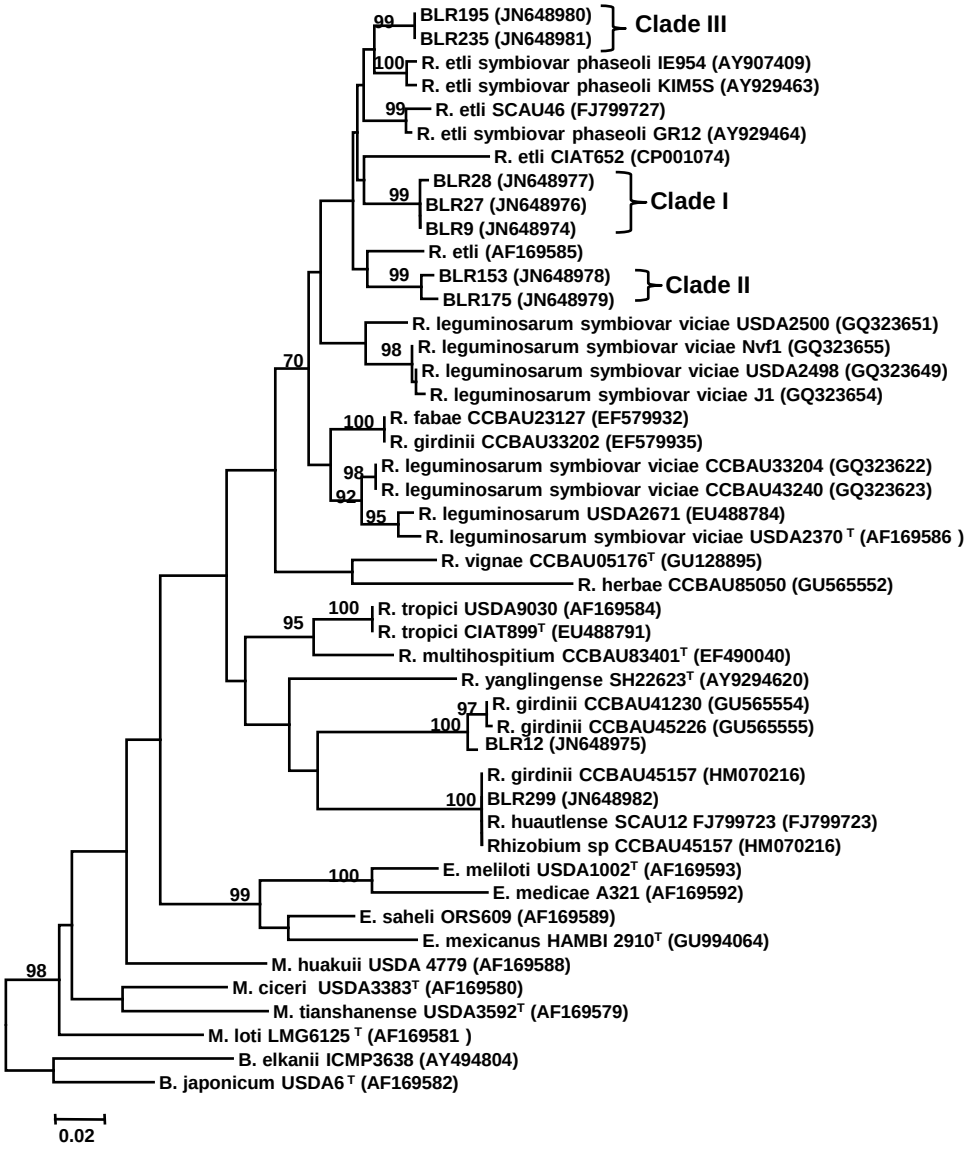


Fig. 12. ML tree based on *glnII* gene partial sequences. Bootstrap values indicated when ≥ 70% (1000 replicates). Abbreviations: BLR: Bangladeshi lentil rhizobia, R: *Rhizobium,* E: *Ensifer,* B: *Bradyrhizobium,* M: *Mesorhizobium.*

Nodulation genes

We amplified three nodulation genes from lentil isolates that were able to form nodules under laboratory conditions, namely *nodA, nodC* and *nodD* (table 6). The nodulation gene sequences showed high similarity to *R. leguminosarum* symbiovar *viciae*. Our phylogenetic analyses based on nodulation genes (*nodA, nodC* and *nodD*) revealed a close relationship to *R. leguminosarum* symbiovar *viciae*, but most of the isolates form distinct clade supported by high bootstrap values (99%, see Figs. 13 – 14 and Fig. 19).

In general, the phylogenetic analyses based on single gene and concatenated sequences of housekeeping genes, resulted in congruent topologies that supported the relationship to *R. etli* but formed three distinct monophyletic groups (Figs. 10 – 12 and Figs. 16 – 17). However, the phylogenetic analyses based on the sequences of the three nodulation genes showed that these three new clades harbor almost identical nodulation genes (Figs. 13 – 14 and Fig. 19) and congruence was absent between chromosomal gene and nodulation gene phylogeny.

DNA fingerprinting

The DNA fingerprints showed three (I – III) different, well-defined banding patterns (see Fig. 15). The NJ analyses based on high resolution ERIC-PCR fingerprints revealed three different clusters. These clusters can be distinguished from each other as well as from *R. etli* and *R. leguminosarum* (Fig. 18a). The principal coordinate analysis (PCoA) showed three non-overlapping clusters and also demonstrated the distinctness of *R. etli* and *R. leguminosarum* with high bootstrap values (Fig. 18b).

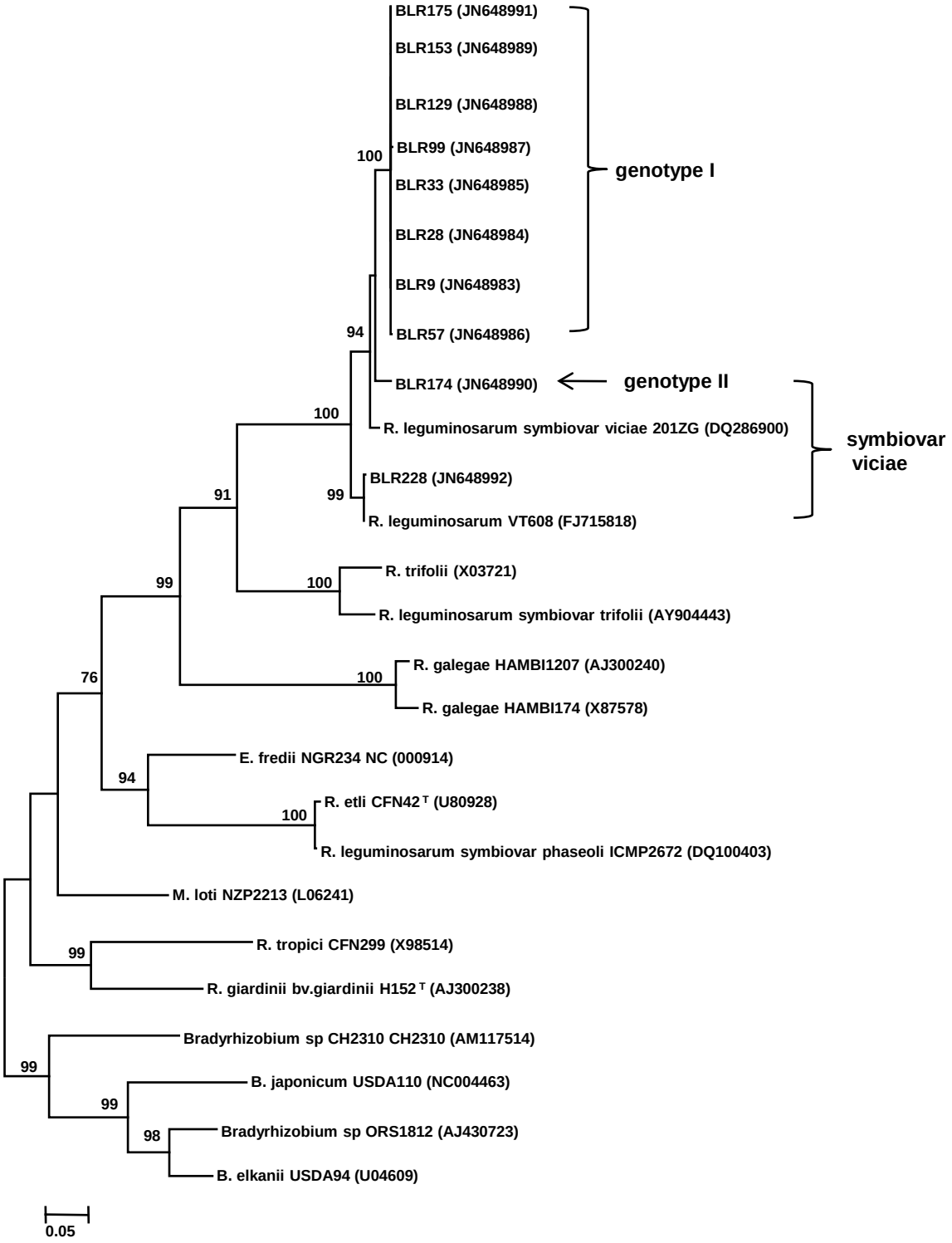


Fig. 13. ML tree based on *nodA* gene partial sequences. Bootstrap values indicated when ≥ 70% (1000 replicates). Abbreviations: BLR: Bangladeshi lentil rhizobia, R: *Rhizobium,* E: *Ensifer,* B: *Bradyrhizobium,* M: *Mesorhizobium.*

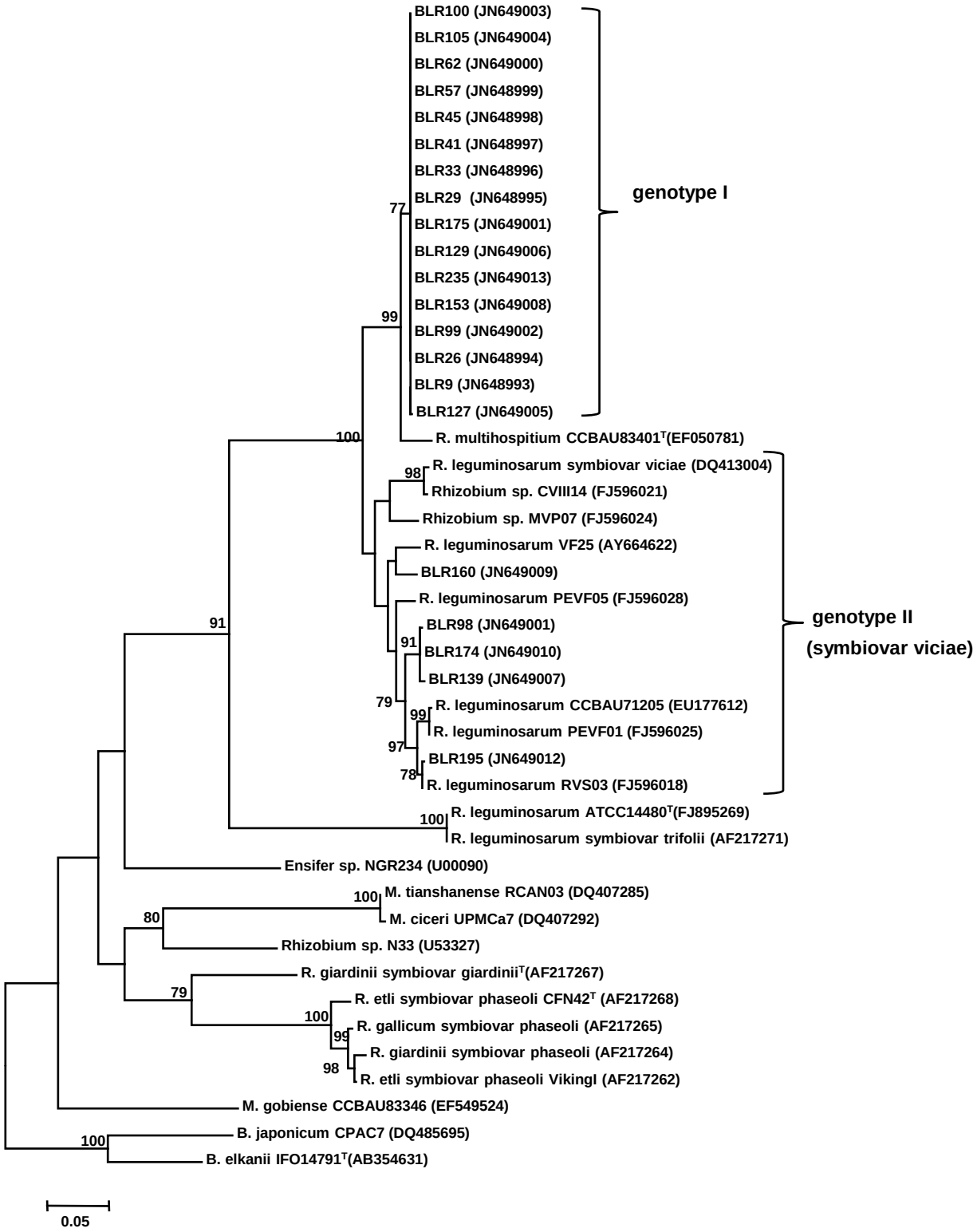


Fig. 14. ML tree based on *nodC* gene partial sequences. Bootstrap values indicated when ≥ 70% (1000 replicates). Abbreviations used: BLR: Bangladeshi lentil rhizobia, R: *Rhizobium,* E: *Ensifer,* B: *Bradyrhizobium,* M: *Mesorhizobium.*

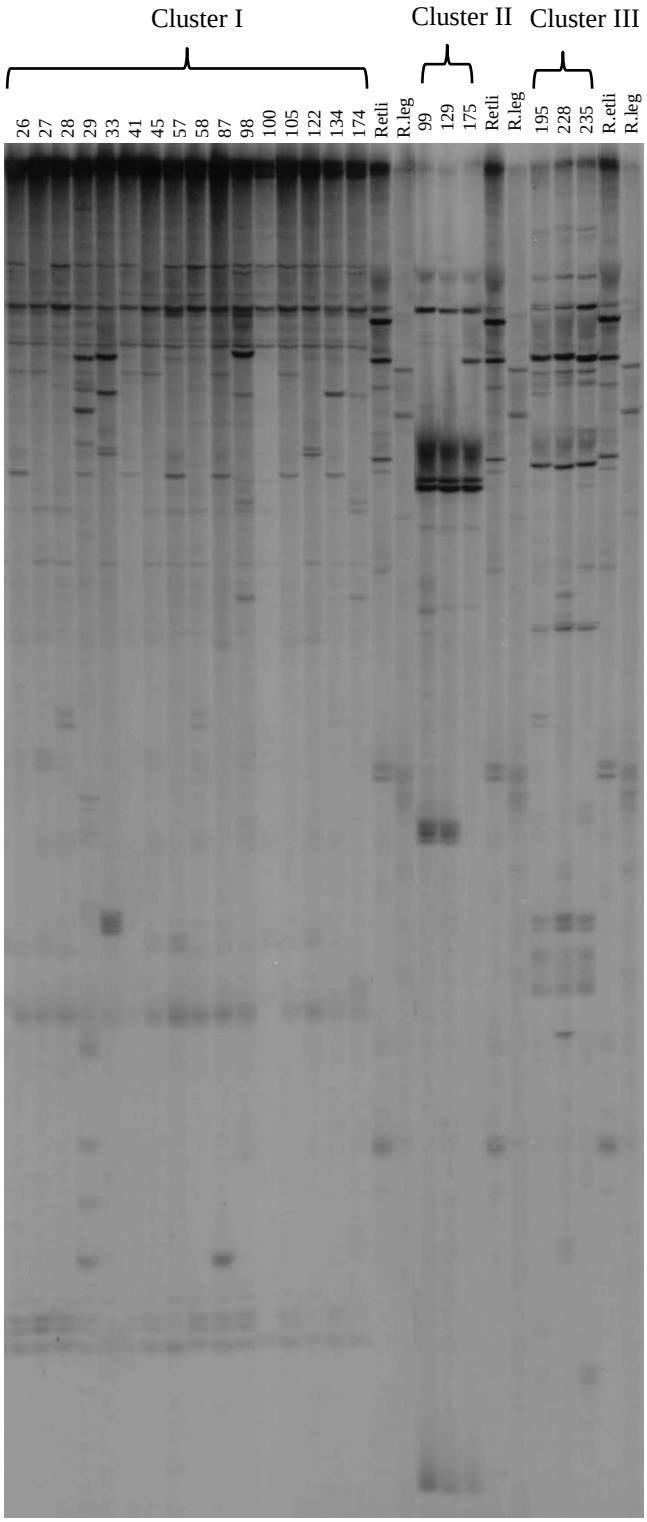


Fig. 15. High resolution ERIC-PCR fingerprint for rhizobial isolates and closely related species including *Rhizobium etli* CFN 42 and *R. leguminosarum* 3841. Abbreviations used: R.etli = *Rhizobium etli*, R.leg = *Rhizobium leguminosarum.*

Phenotypic characterization

The brief phenotypic characteristics of the lentil isolates are listed briefly in table 12. Cells were aerobic, Gram-negative, and rod-shaped. After 48 h of inoculation on YMA media, the diameter of the creamy-white colonies ranged between 1.5 and 2.0 mm, and most of the isolates produced mucilage after 5 days of incubation (table 13). The majority of the isolates grew well on alkaline media (pH 10.0) and showed acidic reactions by producing a yellow coloration on BTB plates. All isolates grew well at 37°C, at pH values of 5.5 – 10, and on media containing both ampicillin (50 g / mL), and kanamycin (10 g / mL), while closely related species like *R. etli* CFN 42 and *R. leguminosarum* 3841 were unable to grow under the same conditions. Like *R. etli*, all isolates (except two) showed resistance to nalidixic acid (up to 40 g / mL) (see table 13).

However, lentil isolates could not grow on LB medium and were sensitive to NaCl; few isolates grew even at 1% NaCl. Out of the different tested antibiotics, tetracycline was the most toxic to lentil isolates; none of the isolates grew on media containing 5 g/mL of tetracycline. The clade III (phylogenetic clade) was more sensitive to ampicillin than clades I and II. None of the isolatess in clade III was able to grow on TY medium containing 100 g / mL ampicillin, while 50% of the mem-bers of clade II and 90% of the members of clade I grew well under the same conditions. Compared to the members of clades I and II, the members of clade III were also more sensitive to kanamycin. In contrast to antibiotic sensitivity, clade III was more resistant to NaCl (1%) and grew better at pH 10.0 than the members of the other two clades (table 13).

Table 12. Brief morpho-physiological characteristics of lentil isolates and their closest relatives (clade wise)

Characteristics (conditions at / in)		Clade I	Clade II	Clade III	*R. etli* (CFN42)	*R. leguminosarum* (3841)
Temperature (°C)	32	+	+	+	+	+
	37	+	+	+	–	–
	40	–	–	–	–	–
	45	–	–	–	–	–
pH	4.5	–	–	–	–	–
	5.5	+	+	+	±	±
	8.2	+	+	+	–	–
	9.0	+	+	+	–	–
	10.0	+(76%)	+ (90%)	+	–	–
NaCl (%)	0.5	+	+	+	–	–
	1.0	–	–	+	–	–
	1.5	–	–	–	–	–
Resistance to antibiotics						
Ampicillin (µg / mL)	50	+	+	+	–	+
	75	+(95%)	+ (50%)	–	–	±
	100	+(90%)	+ (50%)	–	–	–
	125	+(73%)	+ (16%)	–	–	–
	150	+(33%)	+ (16%)	–	–	–
Kanamycin (µg / mL)	10	+	+	+	–	+
	20	+(90%)	+ (83%)	+(66%)	–	±
Nalidixic acid (µg / mL)	10	+(100%)	+(100%)	+(100%)	+	±
	20	+(85%)	+(100%)	+(100%)	+	–
	30	+(85%)	+(100%)	+(100%)	+	–
	40	+(85%)	+(100%)	+(100%)	+	–

3.1.5 Discussion

Rhizobia nodulating lentil are closely related to *Rhizobium leguminosarum* and *Rhizobium etli*

The 16S rRNA gene fulfils all of the requirements for a good phylogenetic marker, such as functional conservation, ubiquitous distribution, conserved size and information content, and the presence of evolutionary conserved regions alongside highly variable structural elements (Ludwig & Klenk, 2005). To date, rhizobial strains isolated from the legume tribe *Vicieae* were all found to be closely related to *R. leguminosarum* (Laguerre *et al.*, 2003; Mutch & Young, 2004; Moschetti *et al.*, 2005; Santillana *et al.*, 2008; Hou *et al.*, 2009; Tian *et al.*, 2010).

In our study, the 16S rRNA gene analyses showed that out of 36 isolates, 30 isolates were closely related to *R. etli*, *R. leguminosarum*, *R. pisi* and *R. fabae*, and only these 30 isolates were able to nodulate lentil under laboratory conditions. The phylogenetic analyses based on 16S rRNA gene sequences indicated that the main rhizobial isoltes nodulating lentil in Bangladesh may be related to *R. etli*. Other isolates that were not able to form nodules under laboratory conditions showed phylogenetic relationships to different rhizobial species such as *R. huautlense*, *R. giardini* and *Ensifer* (*Sinorhizobium*) sp. These rhizobia do not form nodules with host plant species within the tribe *Vicieae*, but have the capacity to form nodules with other legumes (Willems, 2006).

Moreover, we also obtained other bacteria that failed to form nodules under laboratory conditions, and these were closely related to *R. radiobacter* (*Agrobacterium* sp.). There are numerous reports (Anyango *et al.*, 1995; Zakhia *et al.*, 2006; Li *et al.*, 2008; Cummings *et al.*, 2009; Ibañez *et al.*, 2009) that this bacterial species is very often recovered from nodules of naturally growing legumes from around the world. Thus, although bacteria may lose their nodulation genes during storage (Ibañez *et al.*, 2009), these results suggest that these non-nodulating isolates may be oppor-tunistic bacteria capable of lentil nodulation under field conditions, or bacteria that coexist within the nodule alongside nodulating rhizobia.

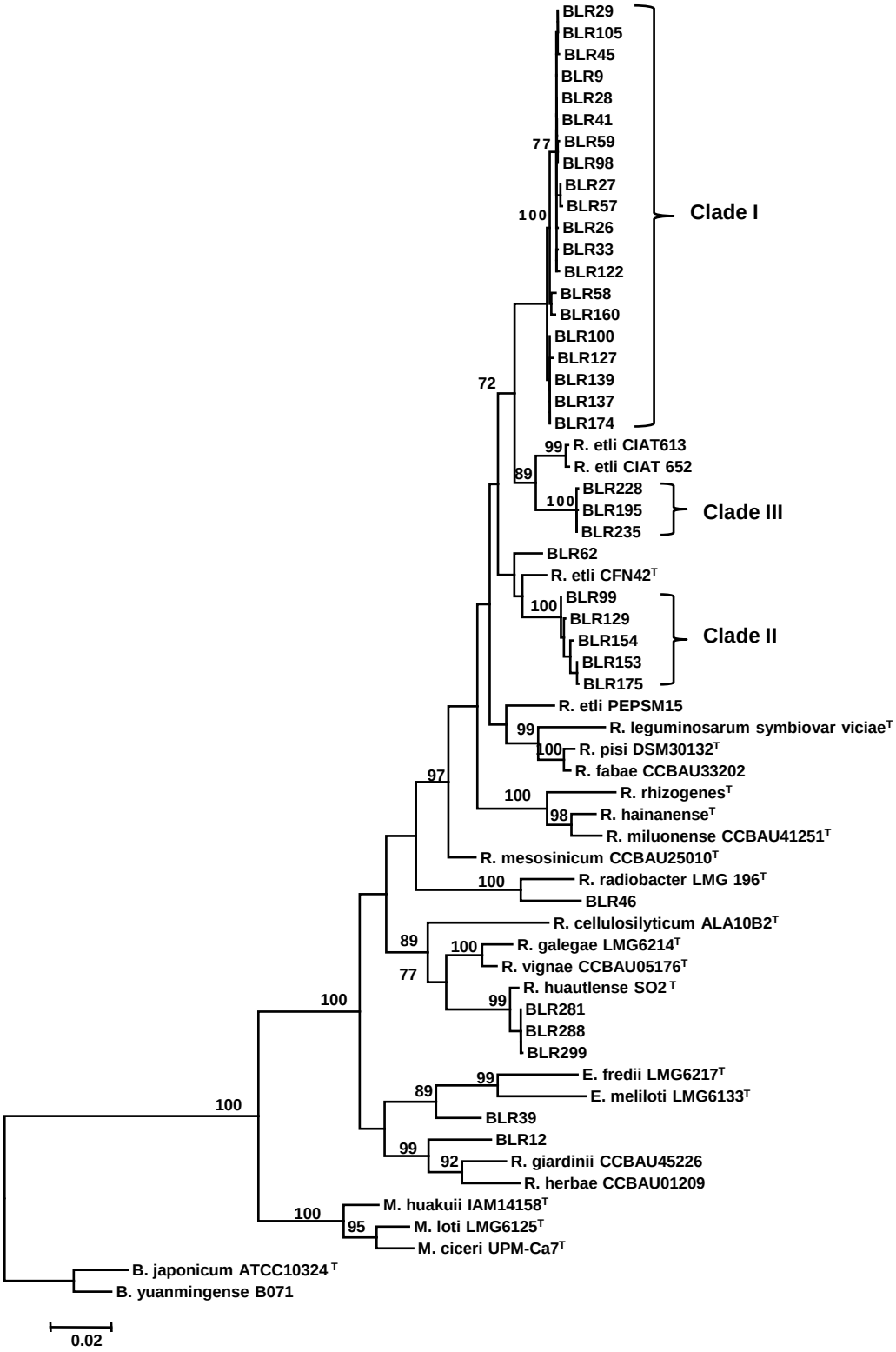


Fig. 16. ML tree based on concatenated partial sequences of 16S, *atpD* and *recA* genes. Bootstrap values indicated when ≥ 70% (1000 replicates). Abbreviations: BLR: Bangladeshi lentil rhizobia, R: *Rhizobium,* E: *Ensifer,* B: *Bradyrhizobium,* M: *Mesorhizobium.*

At least three different rhizobial clades are involved in lentil nodulation

Multilocus sequence analysis (MLSA) is considered to be a better approach for describing the relatedness of different bacterial species by phylogenetic analysis than the analysis of a single locus (Konstantinidis *et al.*, 2006; Martens *et al.*, 2008). Although in a typical MLSA sequences from six to eight genes are used for strain analysis, an accurate phylogenetic reconstruction was obtained by using just three of the best-performing genes (Konstantinidis *et al.*, 2006). MLSA may also uncover horizontal gene transfer or recombination events occurring in one or more lineages showing incongruent phylogenetic signals (Konstantinidis *et al.*, 2006). The housekeeping genes *recA*, *atpD* and *glnII* play vital roles in homologous recombination, ATP synthesis and nitrogen assimilation, respectively. For closely related species, these three protein-coding genes are more informative for phylogenetic analysis than 16S rRNA. Although they are under more stringent functional constraints, they show higher substitution rates than 16S rRNA and have been proven to be phylogenetically informative (Bromofeld *et al.*, 2010 and references therein).

In our study, all ML phylogenetic trees based on individual protein-coding housekeeping genes (*recA, atpD, glnII*), a tree based on the concatenated sequences of the 16S rRNA, *recA* and *atpD* genes, and a tree based on the concatenated 16S rRNA, *recA, atpD* and *glnII* gene sequences showed similar and congruent topologies (Figs.16 – 17). In Bayesian Inference analysis, all housekeeping genes, including 16S rRNA, showed similar congruent results (Figs. Appendix: 1 – 9). By analyzing 16S rRNA, *recA*, *atpD* and ITS nucleotide sequences from several strains isolated from *Vicia faba* and *P. sativum*, Santillana *et al.* (2008) suggested that several species may be described within the *R. leguminosarum–R. etli* group. Our analyses of three protein-coding housekeeping genes (*recA*, *atpD* and *glnII*) supported three distinct clades separate from *R. etli* and *R. leguminosarum* (Figs. 10 – 12), which suggests that lentil-nodulating rhizobia are related to *R. etli* and *R. leguminosarum*, but correspond to species that are genetically distinct from the aforementioned species.

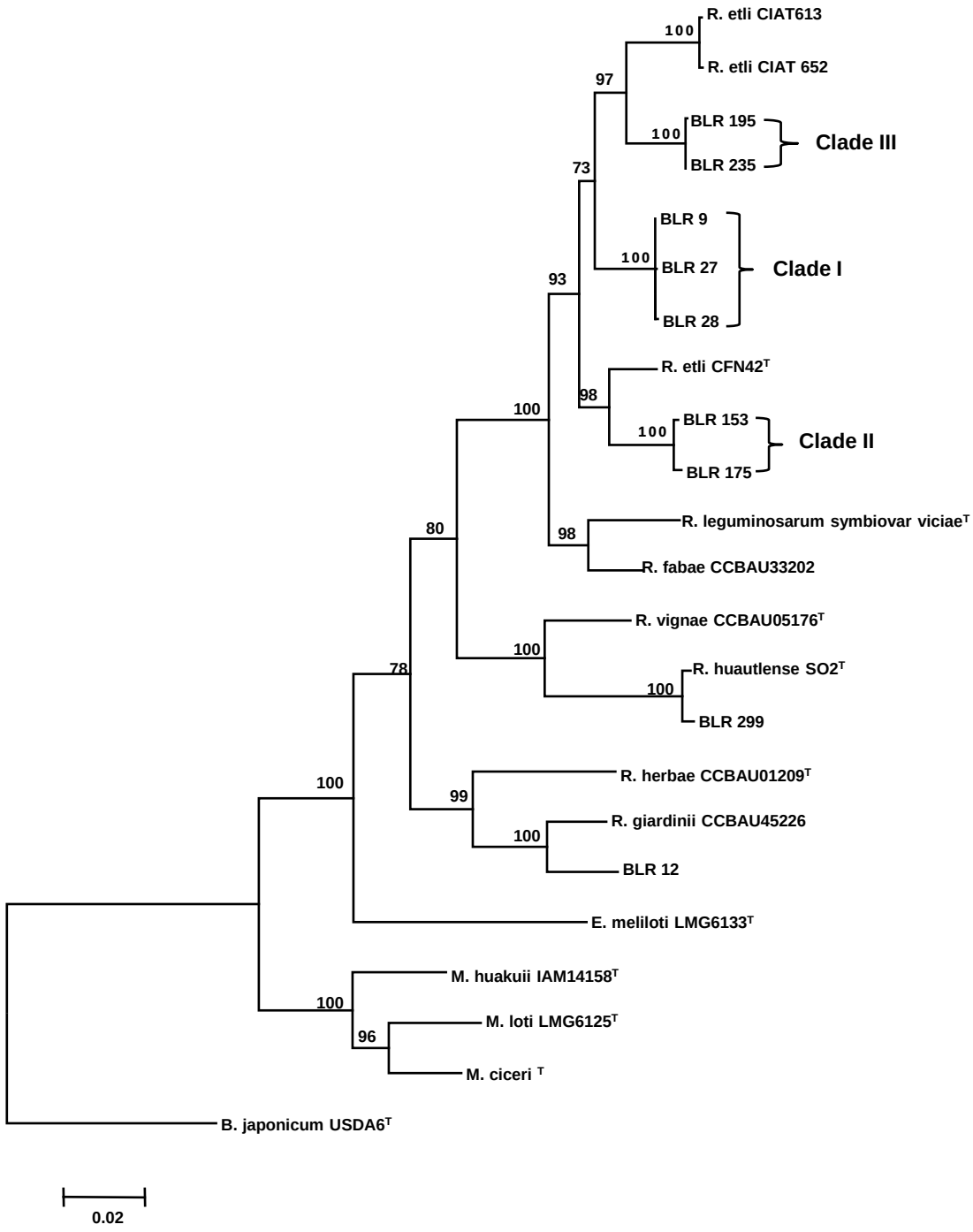


Fig. 17. ML tree based on the concatenated partial sequences of 16S, *atpD*, *recA* and *glnII* genes. Bootstrap values indicated when ≥ 70% (1000 replicates). Abbreviations: BLR: Bangladeshi lentil rhizobia, R: *Rhizobium,* E: *Ensifer,* B: *Bradyrhizobium,* M: *Mesorhizobium.*

Out of the three clades, clades I and II were found to be distributed throughout Bangladesh, while clade III was confined to the southeast region of the country (see localities 1 – 25 in Fig. 4).

The uncorrected genetic distances (similarity levels) based on nucleotide sequences correspond to an important parameter in the description of rhizobial species. Different genes are not under the same selection pressures; hence they show different mutation rates. For example, along with other bacterial genes, the *atpD* and *recA* gene sequences have been us ed (Valverde *et al.*, 2006) to describe a new rhizobial species (*R. lusitanum*), and higher similarity values between *atpD* gene sequences than between *recA* gene sequences have been found within the same strain. Moreover, considering both the *recA* and the *atpD* genes, similarity values of 94% or less may point to separate species status, taking into account homology levels from DNA–DNA hybridization experiments (Valverde *et al.*, 2006; Santillana *et al.*, 2008). In our study three protein-coding genes from lentil rhizobia showed 89.1 – 92.8% of similarity to *R. leguminosarum* and 92.1 – 94.8% similarity to *R. etli*. Although topological incongruence has been detected in some phylogenetic studies carried out on r hizobial species (Eardly *et al.*, 2005), we did in not detect any topological incongruence in our phylogenetic reconstructions when analyzing three protein-coding housekeeping genes and the 16S rRNA gene using both ML and Bayesian analyses. Our phylogenetic reconstruction based on different protein-coding genes supports three distinct sub-lineages (clades I, II and III) in single gene tree and tree from concatenated sequences with high bootstrap values (see Figs. 10 – 12 and Figs. 16 – 17).

Moreover, the genetic similarity values among these clusters, based on t hree different protein-coding genes, range between 91 and 92% for *recA*, 92 and 94% for *glnII* and 89 and 94% for the *atpD* gene (table 11). These genetic similarity levels are comparable to those found between *R. etli* and *R. leguminosarum*, suggesting species status for these clades nodulating lentil in Bangladesh.

DNA fingerprinting distinguishes the new clusters from closely related species

Repetitive Extragenic Palindromic (REP) elements and Enterobacterial Repetitive Intergenic Consensus (ERIC) sequences are widespread in the genome of Gram-negative soil bacteria. In ERIC-PCR, the amplified ERIC-like elements are useful for fingerprinting genera, species, and strains of bacteria. These techniques are con-sidered

powerful tools in bacterial taxonomy and may help in the determination of phylogenetic relationships (Versalobic *et al.*, 1991; De Bruijn, 1992).

In our study, the fingerprint patterns distinguish our Bangladeshi isolates from *R. etli* and *R. leguminosarum*. The NJ bootstrap phylogenetic trees and principal coordinate analyses based on high-resolution ERIC-PCR also strongly support the differentiation of three clusters among the isolatesfrom Bangladesh (Fig. 18a and 18b). Consequently, based on r epetitive intergenic sequences which are distributed across the whole genome, these new clusters differ from *R. etli* and *R. leguminosarum*. These results are congruent with the phylogenetic analyses based on housekeeping gene sequences and provide further evidence for the presence of three new clades of rhizobia capable of nodulating lentil.

Collectively, the evidence obtained from phylogenetic trees based on 16S rRNA genes, three protein-coding housekeeping genes, concatenated sequences, levels of genetic similarity and DNA fingerprinting showed that the new clades described in this study are very close to *R. etli* and *R. leguminosarum*, but may correspond to new bacterial species within the genus *Rhizobium*.

Nucleotide sequences of nodulation genes support a symbiovar similar to *viciae*

The nodulation (Nod) factors determine the legume-rhizobium symbiosis (Laguerre *et al.*, 2001). These genes are essential for nodulation and nitrogen fixation and are either localized on transmissible symbiosis plasmids, e.g. in *Rhizobium* and *Sinorhizobium*, or on the chromo-some, e.g. in most species of *Mesorhizobium* and all *Bradyrhizobium* species (Herrera-Cervera *et al.*, 1998; Laguerre *et al.*, 2003). Because of its regulatory function on the expression of other nodulation genes, the *nodD* gene has been used as a symbiotic marker in the analyses of specificity between rhizobia and host plants such

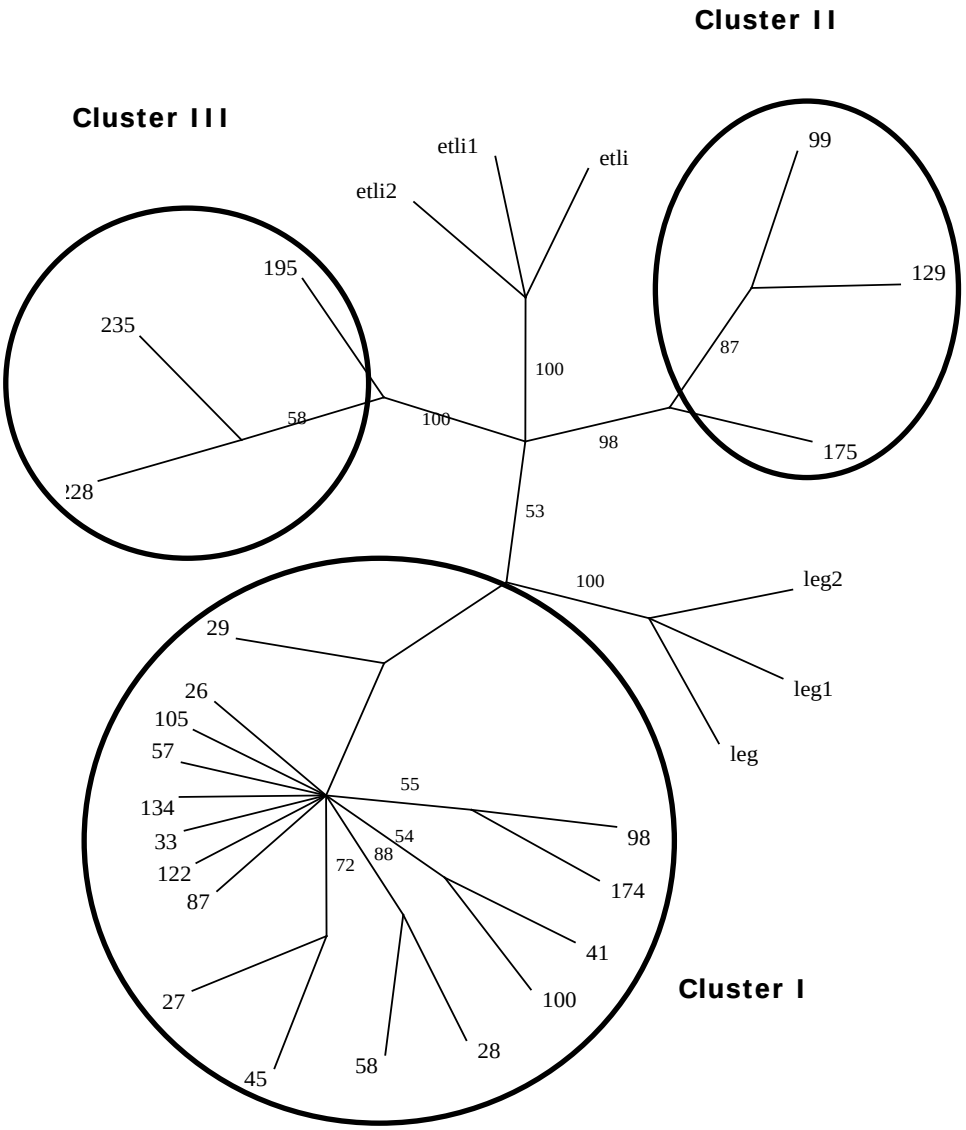


Fig. 18a. The NJ bootstrap trees based on high resolution ERIC-PCR fingerprint.

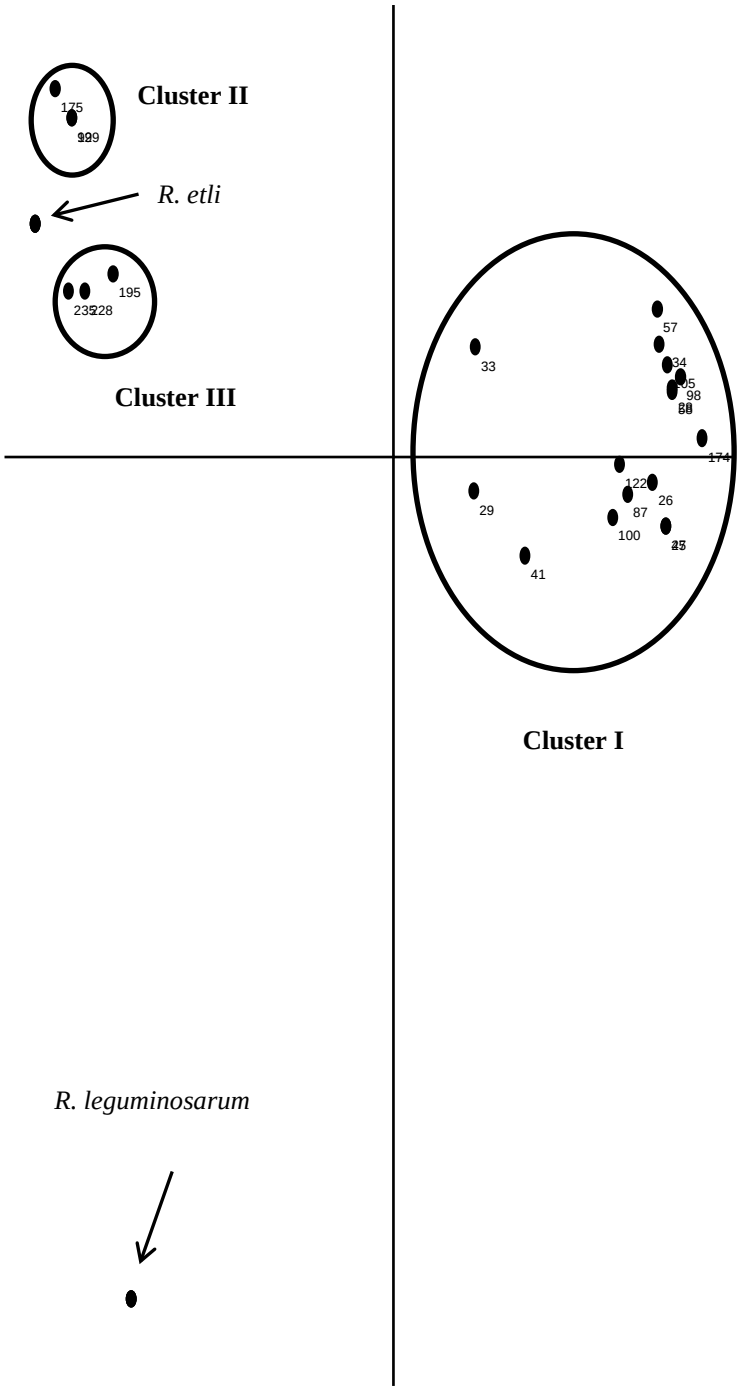


Fig. 18b. PCoA analysis of high resolution ERIC-PCR.

as faba bean (Tian *et al*., 2010; and references therein). Similarly, the *nodC* gene from rhizobia has been widely used to determine the host range of isolates and the degree of host promiscuity (Perret *et al.*, 2000; Laguerre *et al.*, 2001; Iglesias *et al.*, 2007). In this study, three nodulation genes (*nodA, nodC* and *nodD*) were analyzed in order to determine the nature of the symbiotic interaction and the diversity of lentil isolates. Phylogenetic analyses based on three nodulation genes from the lentil isolates revealed high similarities to those of *R. leguminosarum* symbiovar *viciae*, but they still formed a separate clade supported by high bootstrap values (99 – 100%). The symbiovar term has been proposed for rhizobia to describe the adaptive behavior of Rhizobium in legumes, and it should be distinguishable by host ranges as well as by gene sequences (Rogel *et al.*, 2011).

Symbiotic host range and nitrogen fixation ability with various host plants are two important parameters for describing symbiovars (Rogel *et al.*, 2011). However, symbiovars reflect a c omplex phenomenon in rhizobia. For instance, strains of *R. leguminosarum* symbiovar *viciae* may show different host ranges (Rogel *et al.*, 2011 and references therein), and rhizobial strains differ significantly regarding symbiotic effectiveness (Wielbo *et al.*, 2011). *R. leguminosarum* symbiovar *viciae* nodulates lentil effectively (Moawad & Beck, 1991; Tegegn, 2006) and produced effective nodules (pink color) with lentil in our nodulation test. In symbiotic effectiveness test, isolates from three clades (phylogenetic clades) and *R. leguminosarum* symbiovar *viciae* 3841 produced effective nodules and showed higher symbiotic effectiveness than the control treatment. When comparing all BLR isolates versus *R. leguminosarum* symbiovar *viciae* 3841 we found significantly higher nodule dry weights with BLR isolates than with *R. leguminosarum* symbiovar *viciae* 3841, but we did not find the same effect on plant dry weight. We employed only one strain of *R. leguminosarum* symbiovar *viciae*; therefore we can only speculate that the symbiotic performance of the BLR isolates lies within the variability of *R. leguminosarum* symbiovar *viciae*. However, these data should be considered with caution and further research is needed. For example, many more strains of *R. leguminosarum* symbiovar *viciae* and more lentil varieties should be included in order to better understand the symbiovar status of lentil rhizobia.

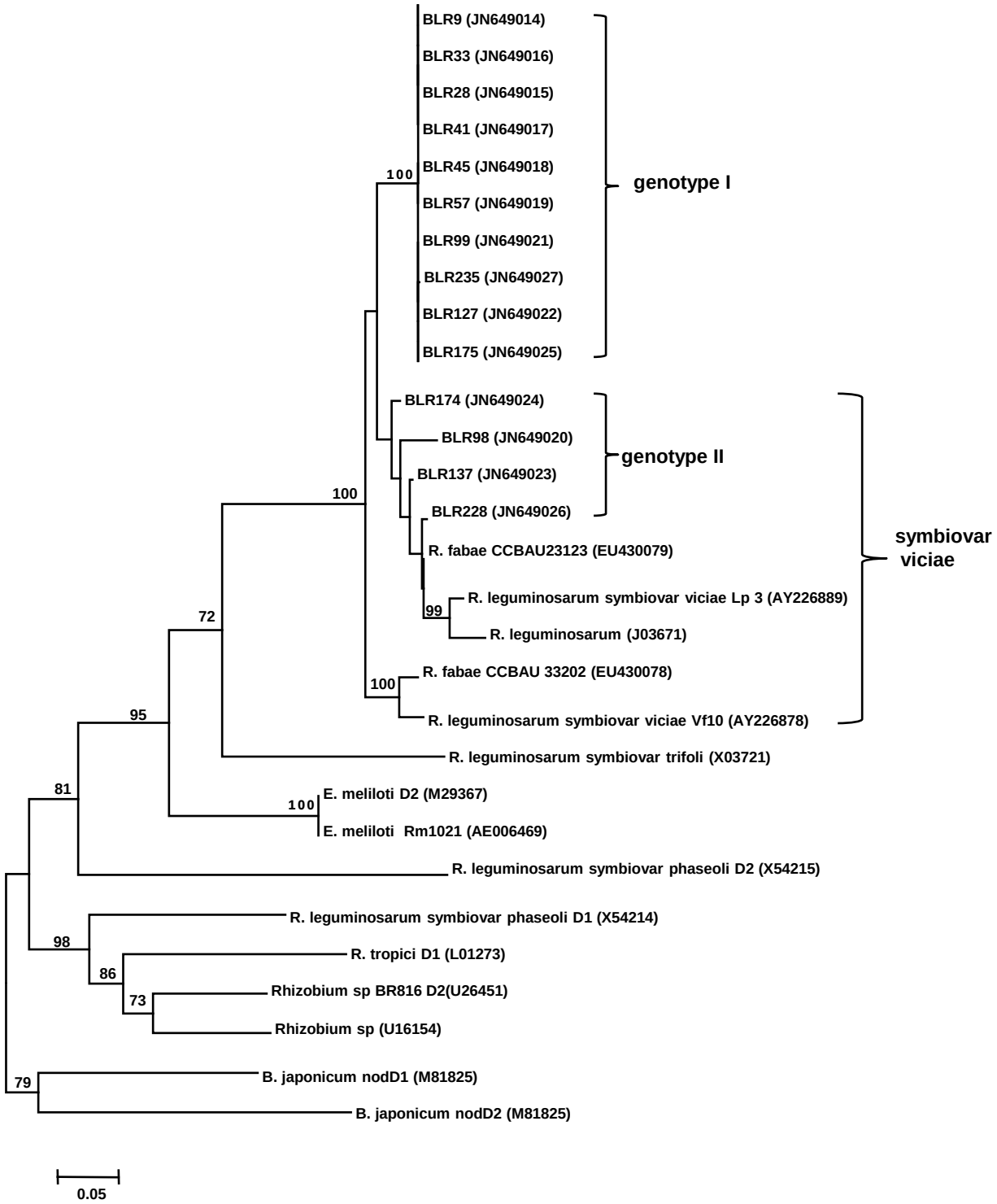


Fig. 19. ML tree based on *nodD* gene partial sequences. Bootstrap values indicated when ≥ 70% (1000 replicates). Abbreviations: BLR: Bangladeshi lentil rhizobia, R: *Rhizobium,* E: *Ensifer,* B: *Bradyrhizobium,* M: *Mesorhizobium.*

In bacteria, the chromosomal genetic background and the plasmid-borne genetic background are often not correlated due to horizontal transfer of plasmid-borne genes (Ochman *et al.*, 2000). Moreover, different rhizobial species can share similar symbiotic genes, and different symbiosis genes can be harbored by similar genomic backgrounds (Laguerre *et al.*, 1996; 2001; Han *et al.*, 2010; Degefu *et al.*, 2011). Thus, discordance may exist between genealogies based on nod and housekeeping genes, and this phenomenon may determine the levels of diversification and structure of natural populations of rhizobia (Young, 1996; Laguerre *et al.*, 2001; Vinuesa *et al.*, 2005). Because of the phylogenetic incongruence between housekeeping genes and symbiotic genes, our results indicate that the symbiotic genes in lentil rhizobial isolates may have evolved independently (Zhang *et al.*, 2001) or evolved from other lineages, probably from *R. leguminosarum* symbiovar *viciae*. Our nodulation and cross-inoculation assays also showed that all isolates behaved similarly to *R. leguminosarum* and were able to nodulate different cultivars of lentil as well as *L. sativus* and *P. sativum*. However, our bacterial isolates related to *R. huautlense*, *R. giardini* and *Ensifer* sp. were unable to form nodules under laboratory conditions, and we were not able to amplify the corresponding nodulation genes using the same primers.

Lentil symbionts are phenotypically different from closely related species

Phenotypic characteristics such as growth at different temperatures, pH values, and sensitivity to different antibiotics allow the distinction between different species of rhizobia. For instance, *R. etli* and *R. leguminosarum* are closely related species, but they differ in terms of phenotypic characteristics (Eardly *et al.*, 2005). Similarly, species of bean-nodulating rhizobia (*R. giardinii* and *R. gallicum*) are phenotypically different from previously described *Rhizobium* species (Eardly *et al.*, 2005). We performed several phenotypic analyses on lentil isolates that showed that lentil isolates were different from closely related species (*R. etli* and *R. leguminosarum*). The isolates grew well on YEMA and TY media at temperatures up to 37°C, at pH 5.5 – 10.0, and at 0.5% NaCl, and hence are more stress-tolerant than closely related species. The colonies showed acidic reactions by producing a yellow coloration on YEMA medium containing BTB, which identified them as fast-growing rhizobia. All isolates showed resistance to ampicillin (50 g / mL), kanamycin (10 g / mL), nalidixic acid (up to 40 g / mL) and sensitivity to tetracycline (5 g / mL). Like *R. etli*, all BLR isolates are nalidixic acid resistant (Segovia *et al.*, 1993). Among the three clades (phylogenetic clades), clade III is more salt tolerant

(1%), but less resistant to ampicillin and kanamycin than the other two clades. Although lentil-nodulating rhizobia could be distinguished from the closely related *R. etli* and *R. leguminosarum* by various phenotypic characteristics, we did not detect clear phenotypic differences between the three clades, except for salt tolerance.

Origin of rhizobia nodulating lentil in Bangladesh

The members of the legume tribe *Vicieae* (*Vicia*, *Pisum*, *Lens* and *Lathyrus*) are commonly nodulated by *R. leguminosarum* symbiovar *viciae*, and this symbiosis has been found in different geographical areas (Laguerre *et al.*, 2003; Santillana *et al.*, 2008; Hou *et al.*, 2009; Tian *et al.*, 2010). The dissemination of nodulating bacteria may be enhanced by the export of plants and seeds (Aguilar *et al.*, 2004). Phylogenetic evidence shows a common origin of *R. leguminosarum* strains that nodulate *Vicia* sp., which might have been dispersed from Europe to several other continents (Álvarez-Martinez *et al.*, 2009).

R. etli symbiovar *phaseoli* was originally described as exclusively nodulating and fixing nitrogen with *Phaseolus vulgaris* (Segovia *et al.*, 1993). In Europe, other species such as *R. gallicum* and *R. giardinii* have been also reported to nodulate *P. vulgaris* (Amarger *et al.*, 1997). Based on the high similarity between symbiotic genes and their co-existence in Europe, *R. etli* symbiovar *phaseoli* strains probably donated symbiotic plasmids to *R. leguminosarum* symbiovar *phaseoli*, and these strains may also have transferred symbiotic plasmids to *R. gallicum* and *R. giardinii* (Laguerre *et al.*, 2001; and references therein).

Table 13. Phenotypic characteristics of lentil-nodulating rhizobial isolates from Bangladesh

Isolates	pH Tolerance				NaCl Tolerance (%)			Temperature Tolerance (^{O}C)			Antibiotic sensitivity (µg/mL)											Colony (mm)
											Ampicillin					Kanam.		Nalidixic acid				
	4.5	5.5	8.2	10	0.5	1.0	1.5	32	37	40	50	75	100	125	150	10	20	10	20	30	40	
BLR9	–	+	+	–	±	–	–	+	+	–	+	–	–	–	–	+	+	+	–	–	–	2.0
BLR26	–	+	+	+	–	–	–	+	+	–	+	+	+	+	+	+	–	+	+	+	+	1.5 – 20
BLR27	–	+	+	–	±	–	–	+	+	–	+	+	+	+	+	+	+	+	+	+	+	2.0
BLR28	–	+	+	+	–	–	–	+	+	–	+	+	+	+	+	+	+	+	+	+	+	2.0
BLR29	–	+	+	+	–	±	–	+	±	±	+	+	+	–	+	+	–	+	+	+	+	1.5 – 20
BLR33	–	+	+	–	–	–	–	+	+	–	+	+	+	+	–	+	+	+	+	+	+	1.5 – 20
BLR41	–	+	+	–	–	–	–	+	+	–	+	+	+	+	–	+	+	+	–	–	–	1.5 – 20
BLR45	–	+	+	–	–	–	–	+	+	–	+	+	+	–	–	+	+	+	+	+	+	1.5 – 20
BLR57	–	+	+	+	–	–	–	+	+	–	+	+	+	+	±	+	+	+	+	+	+	2.0
BLR58	–	+	+	+	±	–	–	+	+	–	+	+	+	+	–	+	+	+	+	+	+	2.0
BLR59	–	+	+	–	–	–	–	+	+	–	+	+	+	+	+	+	+	+	+	+	+	1.5 – 20
BLR62	–	+	+	+	±	–	–	+	+	–	+	+	+	+	+	+	+	+	+	+	+	2.0
BLR87	–	+	+	–	–	–	–	+	+	–	+	+	+	+	+	+	+	+	+	+	+	1.5 – 20
BLR98	–	+	+	–	–	–	–	+	+	–	+	+	+	–	–	+	+	+	+	+	+	1.5 – 20
BLR99	–	+	+	–	–	–	–	+	+	–	+	+	±	–	–	+	–	+	+	–	–	2.0
BLR100	–	+	+	–	–	–	–	+	+	–	+	+	+	+	±	+	+	+	+	+	+	1.5 – 20
BLR105	–	+	+	–	±	–	–	+	+	–	+	+	–	–	–	+	+	+	+	+	+	1.5 – 20
BLR122	–	+	+	+	+	–	–	+	+	–	+	+	+	+	+	+	+	+	+	+	+	2.0
BLR127	–	+	+	–	–	–	–	+	+	–	+	+	+	+	+	+	+	+	+	+	+	1.5 – 20
BLR129	–	+	+	+	+	–	–	+	+	±	+	+	+	–	–	+	+	+	+	+	+	2.0
BLR137	–	+	+	–	–	–	–	+	+	–	+	+	+	–	±	+	+	+	+	+	+	1.5 – 20
BLR139	–	+	+	+	–	–	–	+	+	–	+	+	+	+	±	+	+	+	+	+	+	1.5 – 20
BLR153	–	+	+	+	+	–	–	+	+	–	+	–	–	–	–	+	+	+	+	+	+	2.0
BLR154	–	+	+	+	+	–	–	+	+	–	+	–	–	–	–	+	+	+	+	+	+	2.0
BLR160	–	+	+	–	–	–	–	+	+	–	+	+	+	+	±	+	+	+	–	–	–	1.5 – 20
BLR174	–	+	+	+	–	–	–	+	+	–	+	+	+	+	±	+	+	+	+	+	+	1.5 – 20
BLR175	–	+	+	–	+	–	–	+	+	–	+	–	–	–	–	+	+	+	+	+	+	2.0
BLR195	–	+	+	+	+	–	–	+	+	–	+	–	–	–	–	+	–	+	+	+	+	1.5 – 20
BLR228	–	+	+	±	+	+	–	+	+	–	+	–	–	–	–	+	+	+	+	+	+	1.5 – 20
BLR235	–	+	+	+	+	+	–	+	+	–	+	–	–	–	–	+	+	+	+	±	±	1.5 – 20
R.etli	–	±	–	–	–	–	–	+	–	–	–	–	–	–	–	–	–	+	+	±	±	2.0
Rlv3841	–	±	–	–	–	–	–	+	–	–	+	±	–	–	–	+	+	±	–	–	–	1.5 – 20

+ = growth positive, – = no growth, ± = growth was not clear, NB = non nodulating isolates were not considered for phenotypic characterization, at kanamycin 200 µg / mL and neomycin 100 µg / mL, LB medium : no isolates and type strains showed growth, at 4^{O}C and 32 ^{O}C all isolates and type strains were positive

Seeds are known carriers of bacteria. For example, the testa of the seeds of *P. vulgaris* has been shown to carry rhizobia and play an important role in the dispersal of R. etli in different geographical regions (Perez-Ramírez *et al.*, 1998). The host plant of *R. etli* is *P. vulgaris*. The *P. vulgaris* is neither a well-known crop nor widely cultivated in Bangladesh, leaving a low probability of transferring *R. etli* or close relatives from Europe to Bangladesh via *P. vulgaris*. In contrast, the lentil has been cultivated in Bangladesh since antiquity, and the isolates collected in this study were from field-grown lentil plants from all over the country. The center of origin for lentil is located in the Near East and Central Asia (Shandu & Singh, 2007; and reference therein) and this crop is not being cultivated on a very large scale in Europe. Among rhizobia, parallel and independent evolution may occur in different locations (Wolde-Meskel *et al.*, 2005) leaving these new lineages as potential endemic species in Bangladesh soil.

We have characterized 30 rhizobial isolates, isolated from lentil grown in Bangladesh, using a polyphasic approach. Our analyses revealed that these lentil isolates are related to *R. etli* and *R. leguminosarum* but differs considerably from the aforementioned species in terms of genetic and phenotypic characteristics. This conclusion is underpinned by phenotypic characterization, DNA fingerprinting using high resolution ERIC-PCR, and multilocus sequence analyses (MLSA). Therefore, there are tentatively three clades involved in lentil nodulation in Bangladesh, and they may correspond to novel species within the genus *Rhizobium*.

3.2 Project 2: *Rhizobium leguminosarum* symbiovar *viciae* is the symbiont of lentils in the Middle East and Europe but not in Bangladesh

3.2.1 Abstract

Lentil is the oldest of the crops that have been domesticated in the Fertile Crescent and distributed to other regions during the Bronze Age, making it an ideal model to study the evolution of rhizobia associated with crop legumes. Housekeeping and nodulation genes of lentil nodulating rhizobia from the region where lentil originated (Turkey and Syria) and regions to which lentil was introduced later (Germany and Bangladesh) were analyzed to determine their genetic diversity, population structure and taxonomic position. There are four different lineages of rhizobia associated with lentil nodulation, of which three are new and endemic to Bangladesh, and one lineage is found to be in Mediterranean and Central Europe that belongs to *Rhizobium leguminosarum*. The endemic lentil grex *pilosae* may have played a significant role in the origin of these new lineages in Bangladesh. The availability of *Rhizobium leguminosarum* with lentil at the centre of origin and at countries where lentil was introduced later suggests that *Rhizobium leguminosarum* is the original symbiont of lentil. Lentil seeds might have played a significant role in the initial dispersal of this species within middle-East and on to other countries. Nodulation gene sequences revealed a high similarity to those of symbiovar *viciae*.

Keywords: Rhizobia; legume; lentil greges; speciation; recombination

3.2.2 Introduction

Nitrogen is an essential nutrient for all living organisms and necessary for high crop yield and plant quality in agriculture, but only prokaryotes can convert atmospheric molecular nitrogen into forms that are available to plants. Rhizobia are nitrogen-fixing soil bacteria that are able to enter a mutual symbiosis with leguminous plants in the form of root nodules that fully or partially satisfy the nitrogen demand of the plant. So far, over 90 rhizobial species from 12 genera of α- and β-proteobacteria have been described that can form nitrogen-fixing nodules with legumes (Masson-Boivin *et al.*, 2009; and references therein). The legume-rhizobium symbiosis is a highly specific mutual interrelationship between the two partners. During the infection process, rhizobia produce a number of host-specific factors, and thus it has been assumed that rhizobia have coevolved with their host plants (Perret *et al.*, 2000) and that host association is important for shaping the genetic divergence of nodulation and housekeeping genes in rhizobia (Wernegreen *et al.*, 1997).

R. leguminosarum is a cosmopolitan and well-studied species in the genus *Rhizobium*. The name *R. leguminosarum* was first proposed by Frank (1889) for all nodule-forming bacteria infecting legumes, but many other species were later described. The current species *R. leguminosarum* has three symbiovars that differ in the selectively plant hosts (Jordan, 1984). The host specificity is conferred by a cluster of *Nod* genes that is usually plasmid-encoded (Perret *et al.*, 2000). Commonly, strains in the symbiovar *viciae* nodulate *Lathyrus*, *Pisum*, *Vicia,* and *Lens* in the legume tribe *Vicieae*. Probably due to its saprophytic nature, *R. leguminosarum* symbiovar *viciae* (Rlv) can maintain high population densities (10^4–10^5) in soil, even if the host plant has been absent for a long time (Hirsch, 1996). Like other species of rhizobia, *R. leguminosarum* can be carried on the testa of seeds (Perez-Ramírez *et al.*, 1998), allowing it to disperse to different geographical regions along with the seeds.

Lentil (*Lens culinaris*) is the oldest crop that was domesticated in the Fertile Crescent around 9,000 years ago (Zohary & Hopf, 2000; Toklu *et al.*, 2009; and references therein) and remains an important and popular legume employed worldwide for human and animal nutrition and for soil fertility management (Sonnante *et al.*, 2009; and references therein; Sarker & Erskine, 2006). The region of origin encompasses Southeastern Turkey and Northern Syria, including the sources of the rivers Tigris and Euphratec (Lev-Yadun *et al.*, 2000). After domestication, lentil spread to Cyprus in the Neolithic period (Erskine *et al.*, 1994) and disseminated from Southeastern Europe to

Central Europe around the 5,000 years BC via the Danube. From Europe, it was transported to the Nile Valley and from there to Ethiopia. In Georgia, lentil was propagated during the 5,000 to 4,000 years BC and transported to the Indian sub-continent around 2,500 – 2,000 years BC (Sonnante *et al.*, 2009; and references therein).

Nucleotide sequences of the 16S rRNA genes are widely used genetic markers for bacterial identification and classification (Martens *et al.*, 2008), but suffer from a number of known drawbacks. For a more precise identification and description of closely related bacterial species, including rhizobia, multi-locus sequence analysis (MLSA) using different protein-coding genes has become the preferred method (Ludwig & Klenk, 2005; Konstantinidis *et al.*, 2006; Martens *et al.*, 2008). Furthermore, phylogenies inferred from chromosomal and plasmid-encoded symbiotic genes of rhizobia are frequently found to be incongruent due to the existence of horizontal / lateral transfer of plasmids and plasmid-borne genes (Sprent, 1994; Laguerre *et al.*, 1996; Young & Haukka, 1996). Recombination occurs frequently in bacteria and plays an important role in the evolution of most bacterial species, including rhizobia (Silva *et al.*, 2005; Vinuesa *et al.*, 2005; Maiden, 2006; Bailly *et al.*, 2006; den Bakker *et al.*, 2008; Tian *et al.*, 2012). New genetic material can be rapidly introduced by recombination, allowing faster evolution than simple point mutations (Narra & Ochman, 2006; Redfield, 2001).

The diversity of rhizobia from the tribe *Vicieae*, especially from pea, faba bean and grass pea has been studied previously (Laguerre *et al.*, 1996; Mutch & Young, 2004; Hou *et al.*, 2009; Tian *et al.*, 2010; Risal *et al.*, 2012; and many others). In contrast, there have been relatively few studies on rhizobia that nodulate lentil (Hynes & O'Connell, 1990; Moawad & Beck, 1991; Laguerre *et al.*, 1992; Geniaux & Amargr, 1993; Keatinge *et al*., 1995; Rashid *et al.*, 2009; 2012). By analyzing the nodulating rhizobia of *Vicieae* from different countries it has been concluded that *R. leguminosarum* is the main nodulating species (Tian *et al.*, 2010; and references therein), although a distinct but related species has also been described and named *R. pisi* (Ramírez-Bahena *et al.*, 2008) and, almost simultaneously, *R. fabae* (Tian, *et al.*, 2008). Although, found more than once *R. pisi* has not been seen as the sole or main symbiont in wild peas. By contrast, we found that lentils in Bangladesh were nodulated by three distinct species-level lineages related to *R. etli*, while *R. leguminosarum* was absent (Rashid *et al.*, 2012). It is therefore important to examine lentil symbionts from

other geographic regions in order to establish whether lentils are exceptional in having different symbionts from other legumes of the tribe *Vicieae*.

In this study, lentil-nodulating rhizobia were collected from traditional lentil-growing countries Turkey and Syria, representing the region of original domestication, and from Germany, where lentil have been grown sporadically. We compared DNA sequences of rhizobial isolated from three countries where grex *pilosae* is absent (Barulina, 1930; Sarker & Erskine, 2006) with those of isolated previously (Rashid *et al.*, 2012) from field-grown *pilosae* (Dr. A. R. Sarker; Dr. M. Rahman and Dr. M. A. Samad, personal communication) lentils in Bangladesh. The aim of this study was (1) to explore the genetic diversity and identity of lentil-nodulating rhizobia, (2) to evaluate the levels of genetic diversity and population structure of these bacteria from different geographical locations.

3.2.3 Materials and methods

Soil samples, plant growth and nodule separation

Rhizobia were isolated from nodules of lentils (variety BINA-3, grex *pilosae*) grown in potted soil under glasshouse conditions and from field-grown lentils [variety Anicia, small green lentil, originally imported from France, Mr. Woldemar Mammel, president (Vorsitzender) of the Friends Association (Förderverein), Alblinsen- Förderverein für alte Kulturpflanzen auf der Schwäbischen Alb, personal communication]. Nodules from foeld-grown lentils were collected from Lauterach, Baden-Württemberg, Germany. Soil samples were collected from nine locations in Germany, one in Turkey and two locations in Syria (table 3). All soil samples were collected from cultivated soils, except for one that had collected from a forest in Germany. Soil samples were kept well separated and processed for growing plants under glass house conditions within 3 – 12 days of collection. About 2.5 – 3.0 kg of soil was transferred to a surface-sterilized plastic pot in order to grow lentils. One pot per locality was used to grow 2 – 3 lentil plants for five weeks.

Surface-sterilized (one min in 70% ethanol and 3 – 5 min in 3% NaOCl) and pre-germinated (48 h on 1% water agar) lentil seeds were then placed on potted soil. After germination, a maximum of three plants were grown for five weeks. Plants were irrigated alternately (i.e., water, then N-free seedlings solutions, then water, then N-free seedlings solutions, etc) with sterile water and nitrogen-free seedling solutions when

needed. After five weeks, plants were uprooted carefully, the roots washed with water, dried with tissue paper and then preserved on s ilica gel desicent until further processing.

Bacteria isolation

Methods are available in chapter 2 (materials and methods).

Determination of rhizobial population from collected soil

The number of rhizobial cells in one g of collected soil sample was determined following standard protocols described by Brockwell (1963). Five-fold serial dilution with three replicates for each soil sample and four replicates for each dilution were used for plant inoculation. Seeds were surface-sterilized and germinated on agar plates, and the seedlings were transferred to growth medium as described in chapter 2. Harvested plants were scored for the presence or absence of nodules. The population density of soil samples was determined following the MPN table (Brockwell, 1963; and references therein).

Determination of soil pH

Air-dried samples (200 g soil) were first ground and then sieved (2 mm) to remove large particles. From the sieved sample, 10 g were used to determine the pH. From each locality, the soil pH was measured using 0.01 mM $CaCl_2$ following the protocol ISO 2006 (International standard organization, www.iso.org) with a HANNA pH meter (HI 98150).

Nodulation and cross-inoculation tests

Methods are available in chapter 2 (materials and methods).

DNA isolation, PCR and nucleotide sequencing

Standard methods were followed and descriptions are available in chapter 2.

Phylogenetic analyses

Methods are available in chapter 2.

Population genetic analyses

Details of the methods are available in chapter 2.

Recombination and mutation analyses

Methods are available in chapter 2.

3.2.4 Results

Bacterial isolation, soil pH and rhizobial population density

A total of 98 rhizobial colonies were isolated from lentil nodules representing 12 localities in three countries: Germany (*N* = 78), Turkey (*N* = 12) and Syria (*N* = 8), and different genes were sequenced from 58 isolates (table 1). A single colony was isolated from each of the selected nodules. Each colony was purified by repeated streaking on CRYEMA medium and preserved at −80ºC with 25% glycerol and at 4ºC on agar slants for further study. In addition, we included seven isolates from Bangladesh (Rashid *et al.*, 2012) in the analyses for species delineation and comparison. Isolates numbers and the corresponding localities are documented in table 3.

Soil pH ranged between neutral values of pH 6.5 – 7.4, with the exception of one forest soil sample from Heidebuckelweg, Heidelberg that presented an acidic pH of 4.8 (table 3. Rhizobial population density varied across different localities in Germany, from 114 cells/g soil in Bürstadt (Hessen) to 2.18×10^3 cells / g soil in Ostrach (Baden-Württemberg, table 3).

Nodulation, cross-inoculation and symbiotic effectivity test

The nodulation efficiency test showed that all 98 isolates were able to form nodules with lentil within 3 – 4 weeks after inoculation under growth chamber conditions. In a cross-inoculation test, a set of 30 randomly-selected isolates were able to form nodules with both *Lathyrus sativus* and *Pisum sativum* under the same growth conditions. All isolates produced dark pink nodules and plant leaves were darker green compared to the uninoculated controls, demonstrating that all isolates were symbiotically effective.

Phylogenetic analyses based on nucleotide sequences

We amplified the 16S rRNA gene (about 1.5 kbp length) and obtained sequences of about 1,100 – 1,350 bp from 38 rhizobial isolates originating from three different countries. BLAST searches indicated high similarities (99 – 100%) to *R. leguminosarum* symbiobar *viciae*. Phylogenetic analyses based on 16S rRNA sequences revealed that all isolates from the three different geographical origins were closely related to *R. leguminosarum* and separate from the Bangladeshi isolates (Fig. 20).

Phylogenetic analyses based on concatenated sequence of three housekeeping genes (Fig. 21) and individual housekeeping genes, *recA* (415 bp from 58 isolates), *atpD* (472 bp from 55 isolates) and *glnII* (598 bp from 57 isolates) (Fig. 22 – 24) recovered six sub-lineages (IVa – IVf) with high bootstrap support (70 – 90%). However, the recovered tree topologies differed in phylogenetic analyses, i.e. some isolates changed their positions between different sub-lineages depending on the analyzed gene, revealing phylogenetic incongruence among the loci (table 14 and table 15). Defined by high bootstrap support, two new sub-lineages (IVb and IVd) differ from sub-lineages of *R. leguminosarum* described previously from various hosts and geographic regions (Tian *et al.*, 2010). Phylogenetic analyses and genetic distances of three protein-coding genes (Valverde *et al.,* 2006) with respect to the type strain of *R. leguminosarum* (Fig. 21, Fig. S22 – 24 and table 16) suggest that all isolates belong to *R. leguminosarum*. However, we recovered very similar topologies using different tree reconstruction methods, i.e. NJ, ML and BI (appendix 9 – 16).

Table 14. Phylogenetic incongruence using Shimodaira-Hasegawa (S-H) test

Input data set	**Inferred tree topology data set**		
	recA	***atpD***	***glnII***
recA	NS	+	+
atpD	+	NS	+
glnII	+	+	NS

NS= not significant

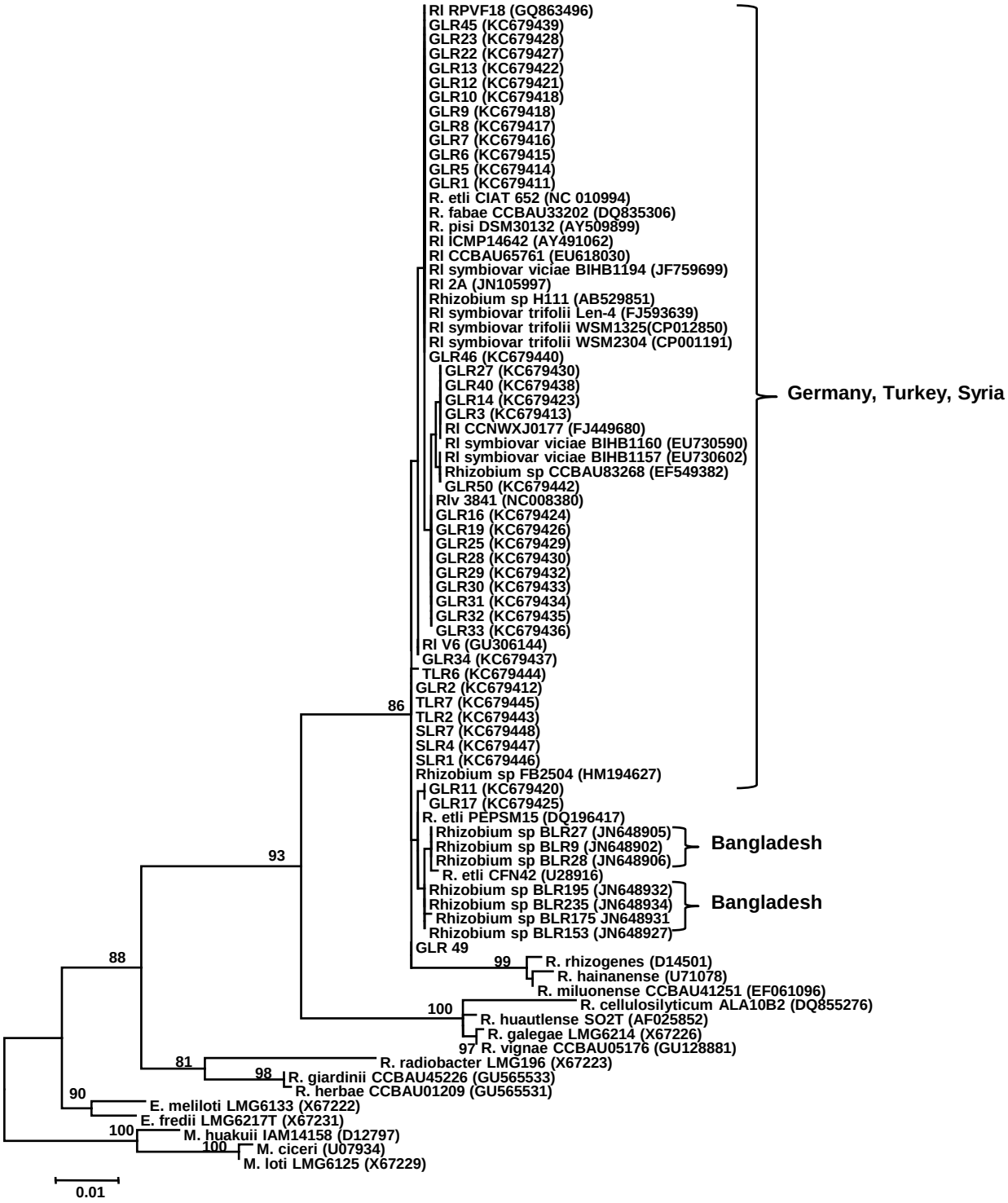


Fig. 20. ML tree from 16S rRNA gene partail sequences. Bootstrap values indicated when ≥ 70% (1000 replicates). Abbreviations: GLR = German lentil rhizobia, TLR = Turkish lentil rhizobia, SLR = Syrian lentil rhizobia, R = *Rhizobium*, Rl = *R. leguminosarum*, E = *Ensifer*, M = *Mesorhizobium* .

Table 15. Isolates under different lineages and sub-lineages in different analyses

Isolates	Phylogenetic analyses				Split analyses (concatenated)	Structure (concatenated)	Recombination (concatenated)		Nodulation genes	
	recA	*atpD*	*gln II*	concatenated (*recA-atpD-glnII*)			with recombination	no recombination	*nodC*	*nodD*
TLR2	a	a	a	a	a	a	a	a	A	C
TLR3	a*	a*	a	a	a	a	a	a	C	
TLR4	a	-	-	-	-	-	-	-	-	
TLR5	a	-	e	a	-	-	-	-	C	
TLR6	a*	a	a	a	a	a	a	a	-	
TLR7	a	a	a	a	a	a**	a	a	A	C
TLR8	-	f	a*	-	-	-	-	-	C	
TLR9	a*	a	a	a	a	a	a	a	C	C
TLR10	a	a	e	a	unique	e**	unique	unique	C	C
TLR11	a	a	a	a	a	a	a	a	C	C
TLR12	a	a	a	a	a	a	a	a	A	C
TLR14	a*	a	a*	unique	unique	e**	unique	unique	C	
SLR1	d	d*	d	d	d	d	d	d	-	B
SLR2	a*	a	a	a	a	a	a	a	unique	-
SLR3	a*	a	a	a	a	a	a	a	E	E
SLR4	a*	a*	a	a	a	a	a	a	E	E
SLR5	a*	a*	a	a	a	a	a	a	E	-
SLR6	a	a*	a	a	a	a	a	a	-	E
SLR7	a*	a*	a	a	a	a	a	a	E	E
SLR8	d	d*	d	d	d	d	d	d	A	A
GLR1	a*	a*	a	a	a	a	a	a	B	B
GLR2	d	d	d	d	d	d	d	d	unique	
GLR3	a*	a*	a	a	a	a	a	a	-	
GLR5	a*	a*	a	a	a	a	a	a	B	
GLR6	unique	f	f	f	f	f	f	f	-	
GLR7	f	f	a*	f	unique	a,f**	unique	b	A	A
GLR8	f	f	f	f	f	f	f	f	D	
GLR9	-	f	f	-	-	-	-	-	-	
GLR10	f	f	f	f	f	f	f	f	unique	
GLR11	d	unique	d	d	d	d	d	d	A	A
GLR12	unique	e	-	-	-	-	-	-	-	
GLR13	a*	a	a	a	a	a	a	a	A	
GLR14	a	-	-	-	-	-	-	-	B	
GLR16	e	a	e*	e	e	e	e	e	D	
GLR17	d	d	d	d	d	d	d	d	B	unique
GLR19	e	e	e	e	e	e	e	e	-	
GLR22	f	f	f	f	f	f	f	f	D	
GLR23	b	b	b	b	b	f**	b	b	D	D
GLR25	e	e	e	e	e	e	e	e	A	
GLR27	a	a*	a	a	a	a	a	a	B	B
GLR28	e	e	e	e	e	e	e	e	-	
GLR29	e	e	e*	e	e	e	e	e	-	
GLR30	e	-	e*	-	-	-	-	-	-	
GLR31	e	e	unique	unique	b	e**	b	e	-	
GLR32	e	unique	e	e	e	e	e	e	-	
GLR33	e	e	b	b	b	b	b	e	A	A
GLR34	d	d*	unique	d	d	d	d	d	A	
GLR40	b	unique	b	b	b	b**	b	e	D	
GLR43	f	f	f	f	f	f	f	f	-	
GLR45	b	b	a	unique	unique	a**	b	b	D	D
GLR46	a*	a*	a	a	a	a	a	a	B	B
GLR49	unique	a*	a	a	a	f**	a	a	A	
GLR50	a	a*	a	a	a	a	a	a	A	A
GLR54	f	f	f	f	f	f	f	f	-	
GLR59	f	f	f	f	f	f	f	f	-	
GLR67	f	f	e	f	f	f**	unique	unique	-	
GLR69	a	e	e	e	c	e	c	c	-	
GLR71	c	c	e	c	c	e	c	c	-	
GLR74	c	c	f	c	c	e**	c	c	-	
GLR79	f	f	f	f	f	f	f	f	-	
BLR9	I	I	I	I	I	I	I	I	I	
BLR27	I	I	I	I	I	I	I	I	-	
BLR28	I	I	I	I	I	I	I	I	-	
BLR153	II	II	II	II	II	II	II	II	I	I
BLR175	II	II	II	II	II	II	II	II	I	I
BLR195	III	III	III	III	III	III	III	III	A	A
BLR235	III	III	III	III	III	III	III	III	-	

Abbreviations: I–III = lineages, a–f = sub-lineages of lineage IV, * = split sub-lineage, **= mixed isolates, - = not sequenced

Species delineation and recombination visualization using neighbor-network analyses

Neighbor-network analyses based on the concatenated data set (*recA-atpD-glnII*) showed a reticulate structure, in which we identified six sub-lineages (Fig. 25a). Two sub-lineages (IVb and IVd) are distinguishable from other described *R. leguminosarum* sub-lineages and strains (Tian *et al.*, 2010), and the sub-lineage IVc was not consistently recovered in all analyses. For example, IVc was recovered in phylogenetic analyses including *recA* and *atpD genes* (Figs: 21 – 22, 23) and TMRCA analyses (Figs. 29a, 29b), but it was not identified in network and STRUCTURE analyses (Figs. 25a, 25b, 26A, 26B). In neighbor-network analysis, by a long edge, lentil isolates from Bangladesh (lineages I, II and III) differed significantly from German, Turkish and Syrian isolates and they did not form any reticulate structure among themselves (Fig. 25a and 25b). The isolates GLR7, GLR45, TLR14 and TLR10, which lay outside the main phylogenetic clusters (Fig. 21), had unique positions in the network with a high level of reticulation, indicating that they are potentially recombinant for one or more genes.

Table 16. Genetic distances among different sub-lineages

Sub-lineages*	**Genes and similarity (%)**			
	recA	***atpD***	***gln II***	**Concatenated *(recA-atpD-glnII)***
****Rlv type vs IVa**	97	93	94	95
Rlv type vs IVb	98	93	94	95
Rlv type vs IVc	97	93	-	96
Rlv type vs IVd	97	93	94	95
Rlv type vs IVe	98	93	98	96
Rlv type vs IVf	99	99	99	99

*All sub-lineages were selected from phylogenetic analyses,
****Rlv = *Rhizobium leguminosarum* symbiovar *viciae* USDA2370^T

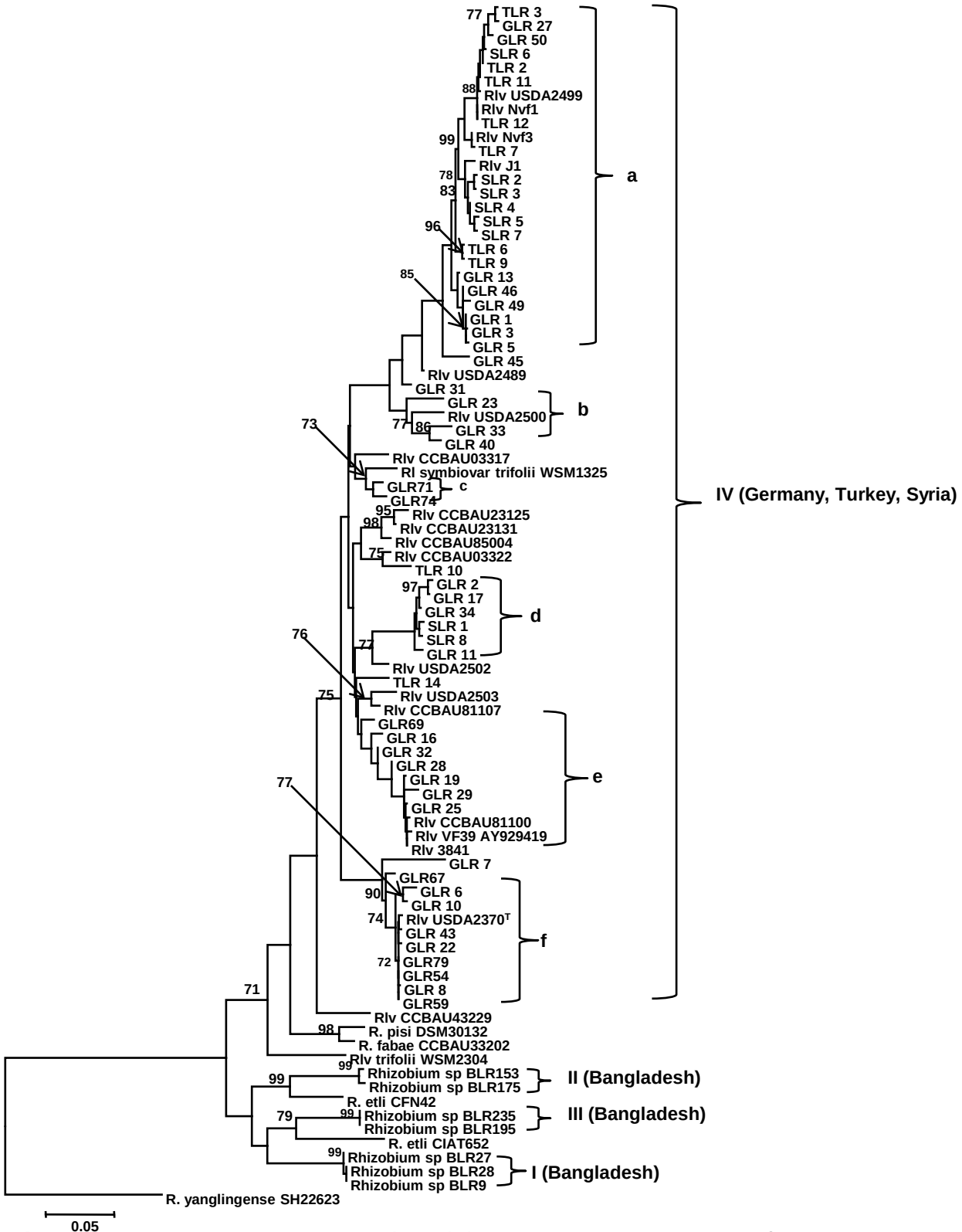


Fig. 21. ML tree from concatenated partial sequences of *recA-atpD-glnII* genes. Bootstrap values indicated when ≥ 70% (1000 replicates). Abbreviations: GLR = German lentil rhizobia, TLR = Turkish lentil rhizobia, SLR = Syrian lentil rhizobia, R = *Rhizobium*, Rl= *Rhizobium leguminosarum,* Rlv = *R. leguminosarum* symbiovar *viciae*, I, II, III, IV = lineages , a – f = sub-lineages within lineage IV.

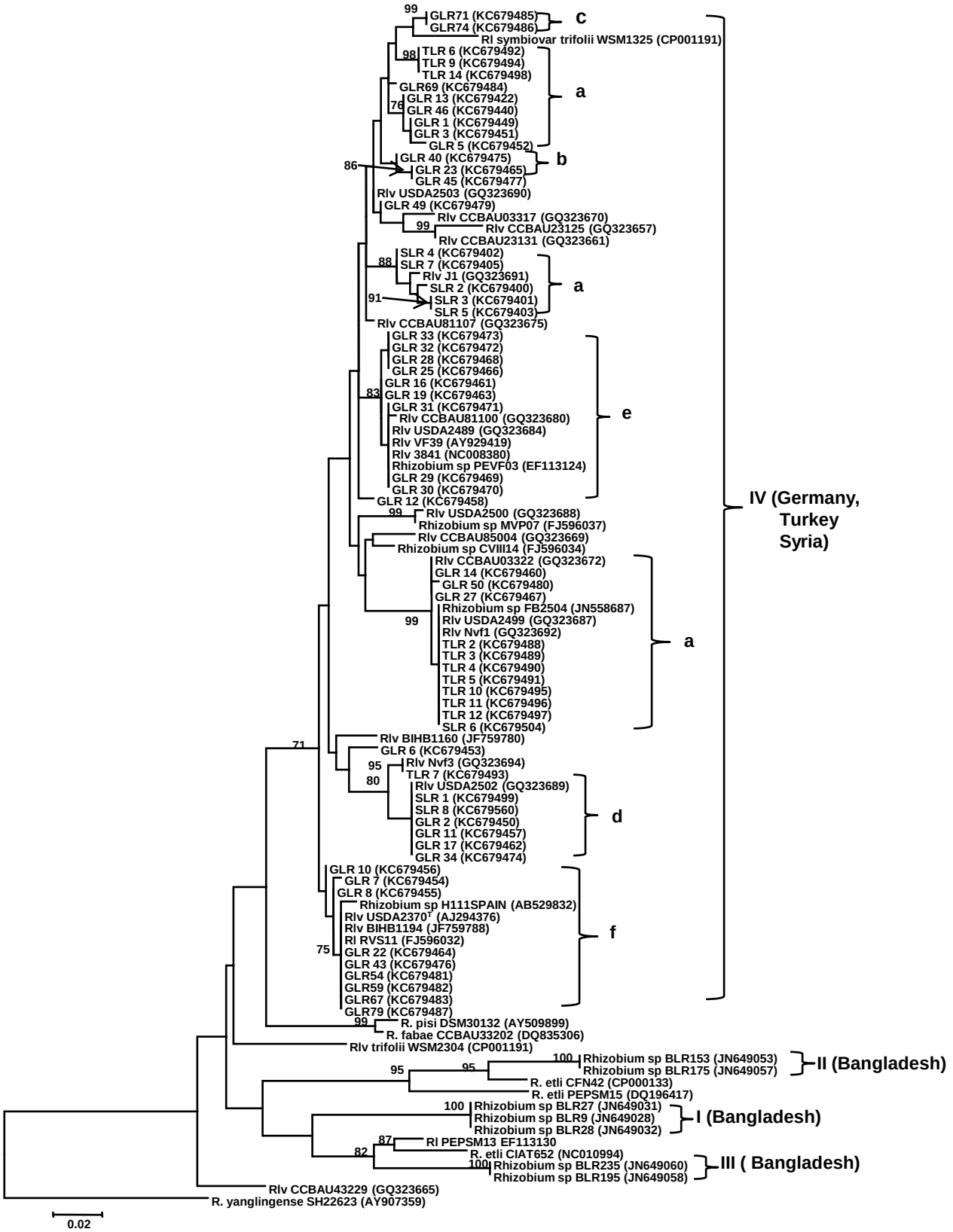


Fig. 22. ML tree from *recA* gene partail sequences. Bootstrap values indicated when ≥ 70% (1000 replicates). Abbreviations: GLR = German lentil rhizobia, TLR = Turkish lentil rhizobia, SLR = Syrian lentil rhizobia, R = *Rhizobium*, Rl= *Rhizobium leguminosarum, Rlv* = *R. leguminosarum* symbiovar *viciae*, I – IV = lineages, a – f = sub-lineages within the lineage IV.

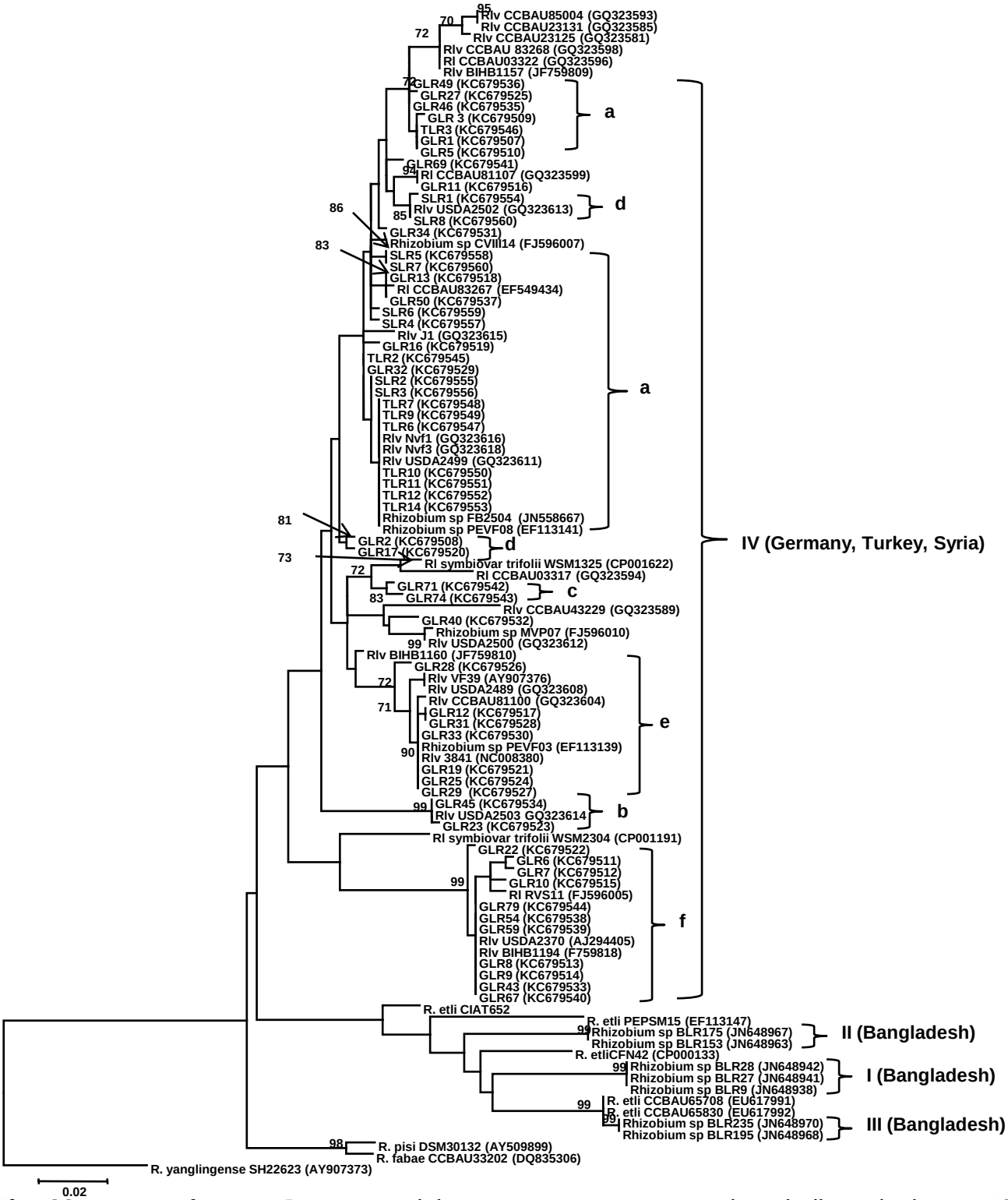


Fig. 23. ML tree from *atpD* gene partial sequences. Bootstrap values indicated when ≥ 70% (1000 replicates). Abbreviations : GLR = Germany lentil rhizobia, TLR = Turkish lentil rhizobia, SLR = Syrian lentil rhizobia, R = *Rhizobium,* Rl = *Rhizobium leguminosarum,*Rlv = *R. leguminosarum* symbiovar *viciae*, I – IV = lineage, a – f = sub-lineages within lineage IV.

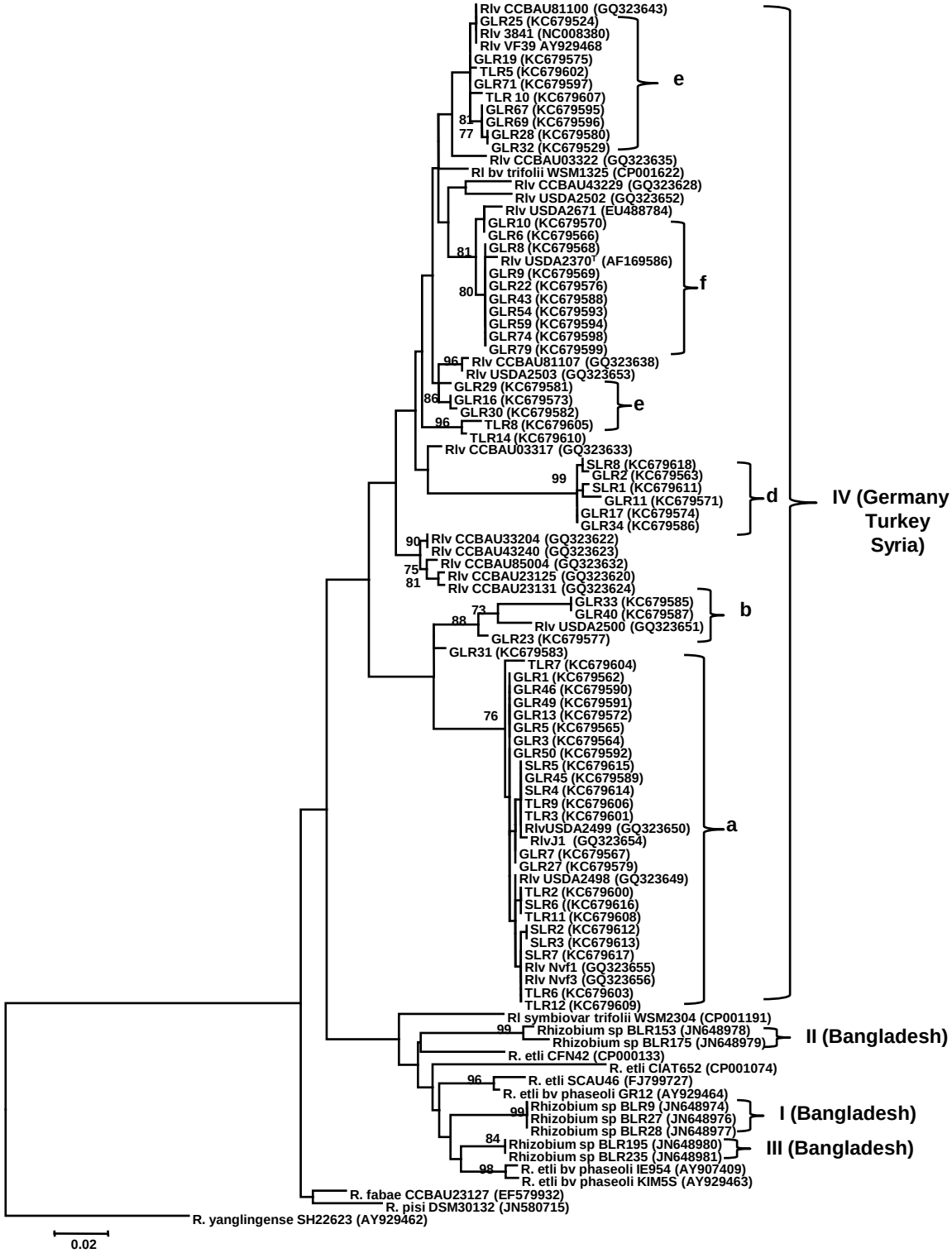


Fig. 24. ML tree from *gln II* gene partial sequences. Bootstrap values indicated when ≥ 70% (1000 replicates). Abbreviations : GLR = German lentil rhizobia, TLR = Turkish lentil rhizobia, SLR = Syrian lentil rhizobia, R = *Rhizobium*, Rl = *Rhizobium leguminosarum*, Rlv = *R. leguminosarum* symbiovar *viciae*, I – IV = lineages, a – f = sub-lineages within lineage IV.

Genetic diversity analyses using STRUCTURE

With five long runs in STRUCTURE, we determined the optimal number of clusters *K* to be 9 and found admixture among populations (Fig. 26A and Fig. 26B). However, we obtained admixed structures in Isolates GLR7, GLR23, GLR31, GLR33, GLR40, GLR45, GLR67, GLR74, TLR7, TLR10 and TLR14. In contrast, the new lineages from Bangladesh did not show any admixture (Fig. 26A and Fig. 26B).

Detection of minimum recombination events and gene flow using DnaSP

Minimum recombination events found in single gene sequences and concatenated sequences of three genes are shown in table 17. Among three housekeeping genes, *recA*, *atpD* and *glnII* reveal 12, 14 and 16 recombination events, respectively, and concatenated data showed 43 recombination events.

Table 17. Minimum recombination events on different genes

Gene	*N*	Gene size (bp)	Rm	Coalescence simulation		
				Rm average	Confidence interval	P< observed Rm
recA	58	415	12	8.65	[5 – 13]	0.945
atpD	55	472	14	3.65	[1 – 6]	1.00
glnII	57	598	16	5.83	[2 – 10]	1.00
concatenated	53	1485	43	17.02	[10 – 24]	1.00

Abbreviations, Rm = minimal intragenic recombination events, N = number of isolates, bp = base pair

High values (Nm = 4.34) of gene flow were found between Turkish and Syrian isolates, along with non-significant K_{ST} values (0.039), while low values (Nm = 1.91) for the same parameters were found between Germany versus Turkey and Syria (table 18).

Differentiation between geographical groups

AMOVA analyses of the concatenated sequence of protein coding genes from isolates from Germany, Turkey and S yria indicated the existence of significant differences between geographical regions (table 18), although the percentage of variation

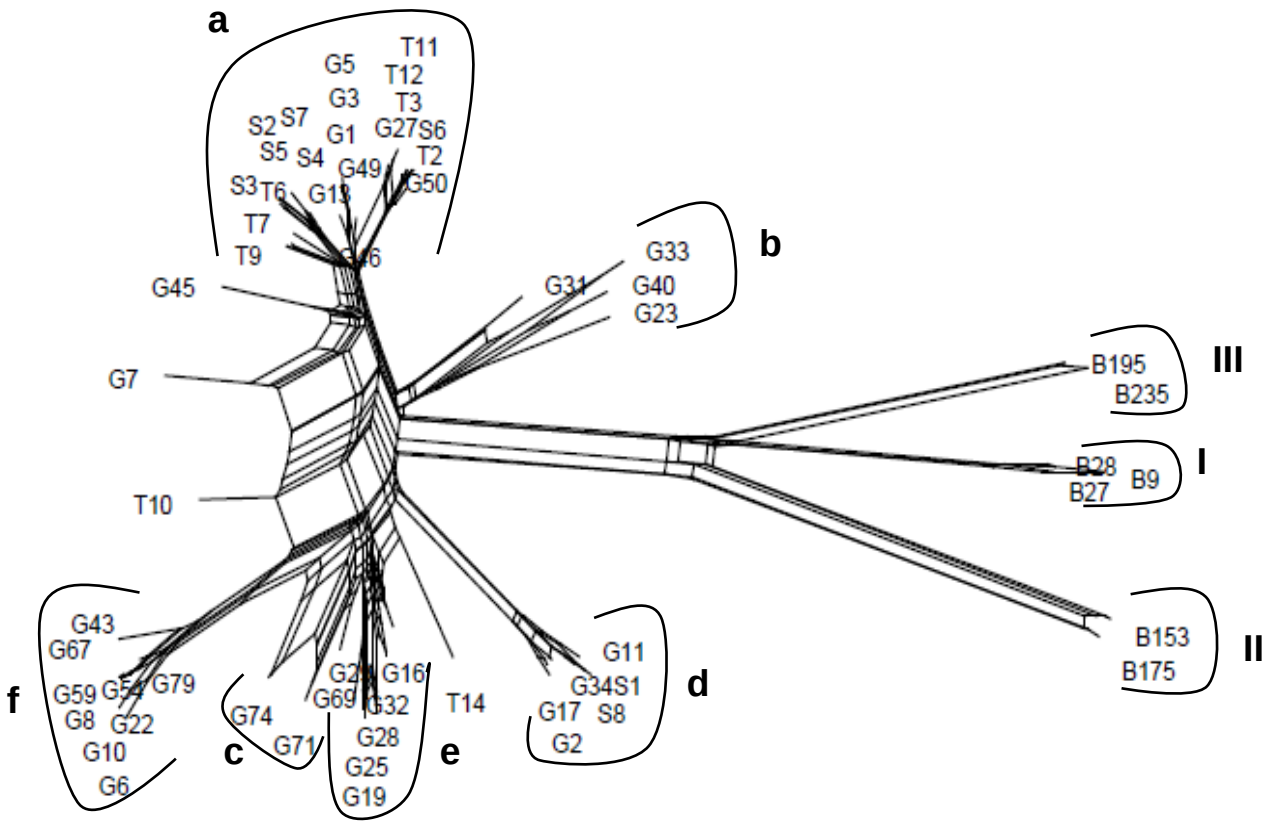
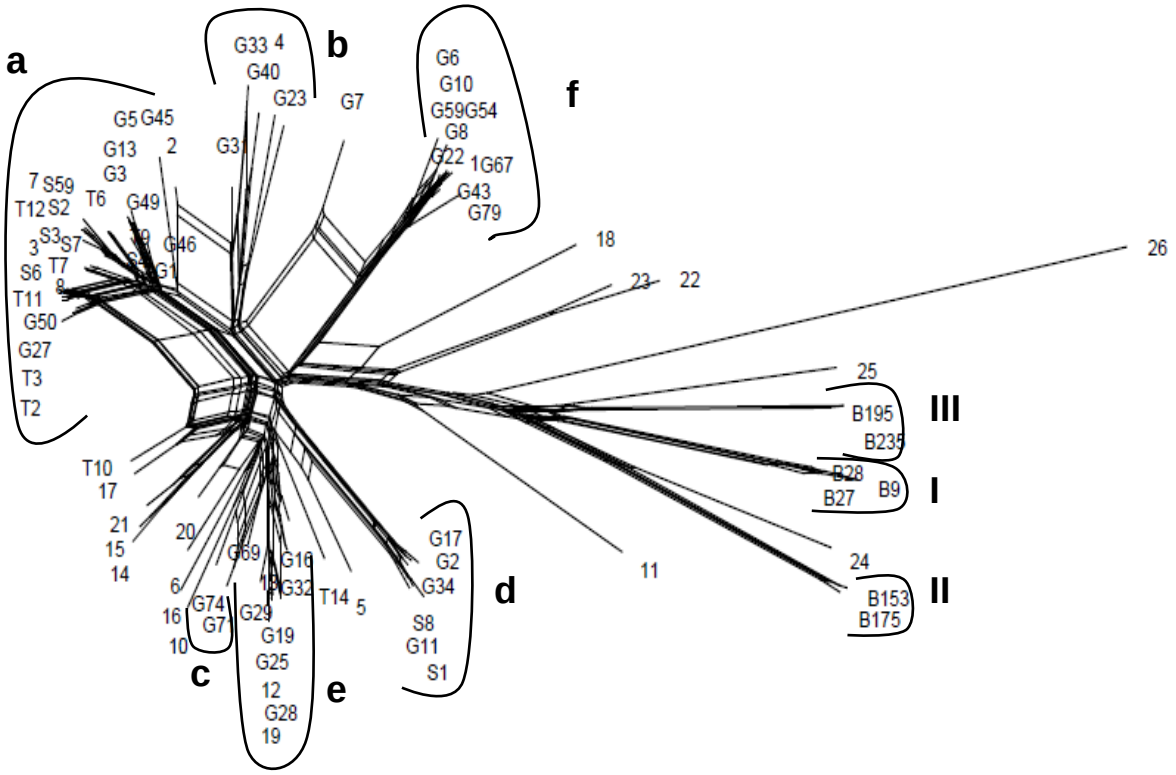


Fig. 25a. Split graph from neighbor-network analyses based on concatenated sequences of *recA-atpD-glnII* genes. Three lineages and six different sub-lineages within the lineage IV were recovered like in ML phylogenetic analyses. Six sub-lineages under the lineages IV are showing clear reticulate structure among themselves indicating the effect of recombination on their genetic structure. Three lineages (I – III) are clearly distinct from the lineage IV recovered from Bangladesh.



B = Bangladeshi lentil rhizobia
G = German lentil rhizobia
T = Turkish lentil rhizobia
S = Syrian lentil rhizobia
Rlv = *Rhizobium leguminosarum* symbiovr *viciae*
I – III = lineages
a – f = sub-lineages within lineage IV

1 = Rlv USDA2370
2 = Rlv USDA2489
3 = Rlv USDA2499
4 = Rlv USDA2500
5 = Rlv USDA2502
6 = Rlv USDA2503
7 = Rlv Nvf1
8 = Rlv Nvf3

9 = Rlv J1
10 = Rlv WSM1325
11 = Rlv WSM2304
12 = Rlv 3841
13 = Rlv VF39
14 = Rlv CCBAU23125
15 = Rlv CCBAU23131
16 = Rlv CCBAU03317

17 = Rlv CCBAU03322
18 = Rlv CCBAU43229
19 = Rlv CCBAU81100
20 = Rlv CCBAU81107
21 = Rlv CCBAU85004
22 = *R. pisi* DSM30132
23 = *R. fabae* CCBAU33202
24 = *R. etli* CFN42
25 = *R. etli* CIAT652
26 = *R. yanglingense* SH22623

Fig. 25b. Split graph from neighbor-network analyses based on concatenated sequences of *recA-atpD-glnII* genes of the studied rhizobial isolates and sequences available in GenBank. Three lineages and six sub-lineages under the lineage IV were recovered like ML phylogenetic analyses. The figure shows relationship of recovered lineages and sub-lineages with previously described *R. leguminosarum*, *R. etli* and other related species. The existence of three new lineages in Bangladesh and two new sub-lineages in Germany, Turkey and Syria are clear in this picture.

Table 18. Hierarchical AMOVA of the genetic structure and gene flow of sub-lineages

Groups	AMOVA						Gene flow	
	Source of variance	d.f.	Sum of squares	Variance component	Variance (%)	Average F_{ST}	*N*m	K_{ST}*
(Syria+ Turkey) vs. Germany	among populations	1	156.22	5.70	18.82	0.18***	1.87	0.04 **
	within populations	51	1253.85	24.58	81.18			
	total	52	1410.07	30.28				
Turkey vs. Germany	among populations	1	116.31	6.30	19.86	0.19***	1.50	0.03 **
	within populations	43	991.44	25.46	80.14			
	total	44	1107.76	30.77				
Syria vs. Germany	among populations	1	96.43	5.36	17.01	0.17***	1.91	0.02 **
	within populations	42	1099.65	26.18	82.99			
	total	43	1196.09	29.85				
Syria vs. Turkey	among populations	1	33.18	1.95	10.50	0.10 ns	4.14	0.04 *
	within populations	15	249.63	16.65	89.50			
	total	16	282.82	18.59				

Abbreviations: *N*m = gene flow, ns = not significant; *, 0.01<P<0.05; **, 0.001<P<0.01; ***, P<0.001

remained low among populations (11 – 20%) compared to the variation found within populations (80 – 90%). However, there were no s ignificant differences between Turkish and Syrian isolates ($P > 0.05$; table 18). Overall, AMOVA analyses showed that German isolates significantly differed from Turkish and Syrian isolates, with Turkish samples differ most (table 18).

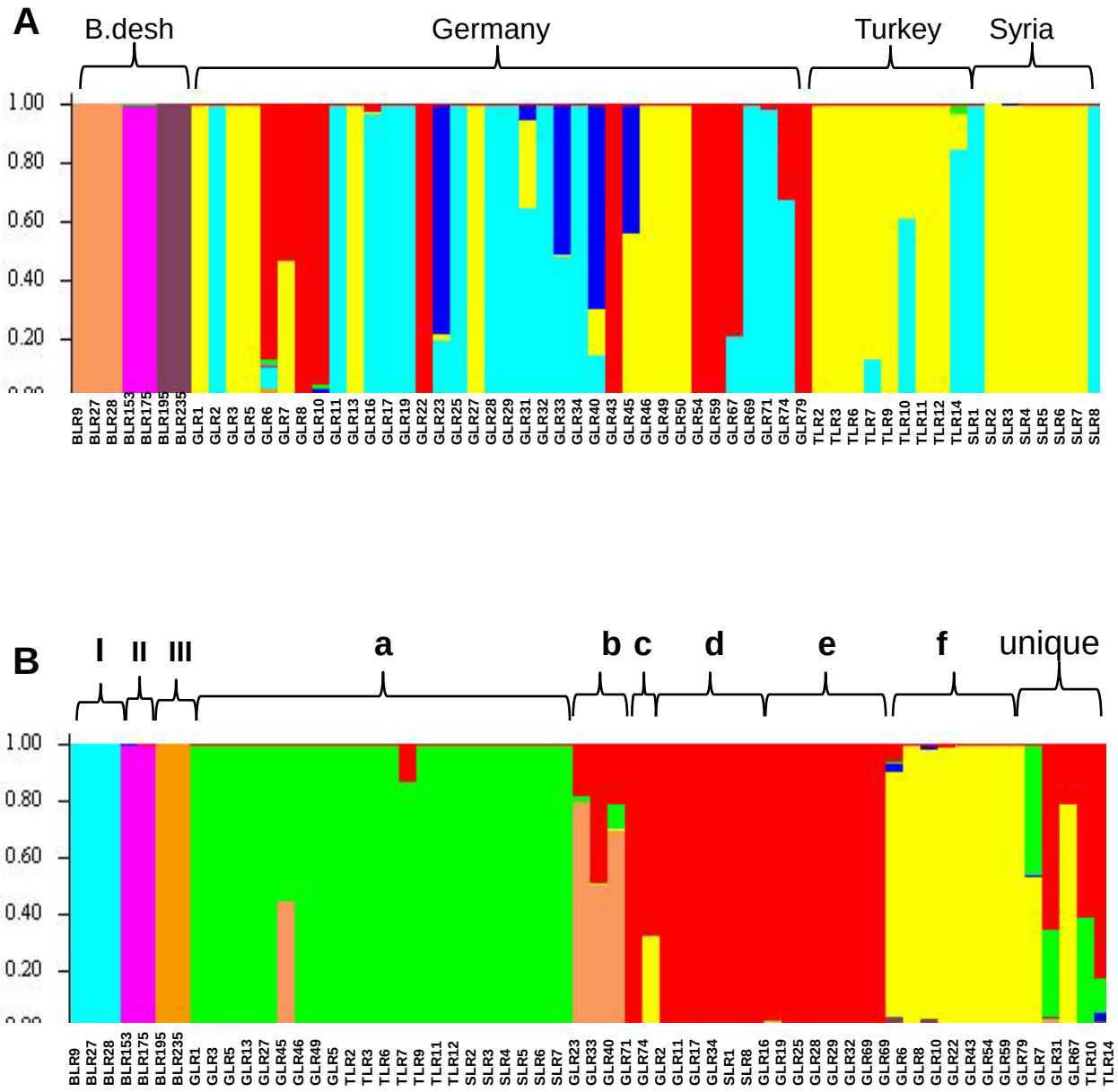


Fig. 26. Ancestry of different lineages and sub-lineages as inferred by STRUCTURE assuming *K* = 9 ancestral populations. Individual isolate is marked at horizontal axis and their corresponding column is filled with color according to the inferred proportion which was inferred from one of the ancestry. Bar plot (A) showing clustering of different group ordered by different countries and bar plot (B) showing clustering of different group arranged by different lineages and sub-lineages.

Relative impact of recombination and point mutation

Using concatenated sequences, we determined the relative effect of recombination versus point mutations. The r/m was 1.52 and the ρ/θ was 0.269, suggesting a greater importance of recombination over mutation for explaining the observed genetic diversity (table 19). Reconstructed phylograms from CLONALFRAME analyses revealed much shorter times to the most recent common ancestor (TMRCA) when considering recombination [TMRCA = 0.220 (0.130 – 0.352), Fig. 29a, table 19] than when assuming no recombination [TMRCA = 1.110 (0.342 – 3.073) Fig. 29b]. Tree topologies also differed between these phylograms.

Table 19. Recombination effect inferred by CLONALFRAME (confidence intervals are shown in parentheses)

Run	R	r/m	ρ/θ	θ
1	26 (12 – 47)	1.479 (0.833 – 2.330)	0.252 (0.112 – 0.471)	109 (71 – 160)
2	31 (14 – 57)	1.611 (0.908 – 2.544)	0.294 (0.125 – 0.550)	109 (72 – 160)
3	26 (12 – 46)	1.490 (0.855 – 2.408)	0.262 (0.121 – 0.502)	101 (67 – 146)
Average	28 (13 – 50)	1.520 (0.865 – 2.428)	0.269 (0.119 – 0.507)	106 (70 – 155)

Abbreviations, R = recombination rate, r/m = relative impact of recombination as compared with point mutation, ρ/ θ = relative frequency of the occurrence of recombination as compared with point mutation, θ = mutational rate

For example, the position of the sub-lineages e and f, and the polytomy including sub-lineages a and f is well resolved assuming no recombination (Fig. 29b). However, we obtained more polytomies in the dendrogram when considering recombination (Fig. 29a).

Symbiotic gene analyses

ML analyses based on nodulation gene sequences recovered five groups (*nodC* from 38 isolates and *nodD* from 24 isolates) from Germany, Turkey and Syria (Fig. 27 and Fig. 28). In *nodC* gene analysis, group A contained isolates from Germany, Turkey and Syria. This group clusters with previously described strains isolated from Peru, Spain, and United Kingdom from different members of the legume tribe *Vicieae*. This group includes the isolate BLR195, which was isolated from lentils in Bangladesh and belongs to lineage III, suggesting a c lear case of horizontal transfer of nodulation genes between lineage III and IV. In same tree (*nodC*) group B and G contained isolates from Germany only, while groups C and E contained isolates from Turkey and Syria,

respectively. Among seven groups (A – E from Germany, Turkey and Syria, and groups I – II from Bangladesh), the group C from Turkey and the group II from Bangladesh differed considerably from existing *nodC* sequences of *R. leguminosarum* symbiovar *viciae* strains. Moreover, two isolates GLR2 and SLR2 showed significant differences from existing strains. In terms of congruence, the nodulation gene trees differed from trees based on chromosomal genes, suggesting horizontal / lateral gene transfer of nodulation genes.

We sequenced *nodD* gene from 24 isolates (11 from Germany, 6 from Turkey, 6 from Syria and one from Bangladesh) to compare with the *nodC* gene analyses to determine whether they were congruent or not. The reconstructed ML tree from the *nodD* gene sequences showed similar tree topologies (except for the isolates GLR17 and SLR 4) to the one based on *nodC* gene sequences (Fig. 28). The exception of GLR17 and SLR4 isolates may arise from internal rearrangement of nod region by recombination. Group A corresponds to the previously described *nodD* type II / *nodD* type g (Laguerre *et al.*, 2003; Mutch & Young, 2004; Tian *et al.*, 2010; and reference therein) from faba bean rhizobia from different geographical locations (Jordan, Spain, Canada and UK). Group B showed similarity with previously described *nodD* type III from faba bean rhizobia and from Europe (France and UK) and China (Tian *et al.*, 2010).

Although Turkish isolates form a separate group (group C) in the *nodC* gene tree, this group showed similarity with a previously described *nodD* type I from the Middle East and China, suggesting rearrangement within the nod region by recombination. Although the *nodC* group from Syria (E) was close to previously described strains, in the *nodD* gene tree this group formed a strong separate group from existing *nodD* groups. The isolate GLR17 had an identical sequence to a distinct strain previously found in France from pea rhizobia. However, Bangladeshi isolates formed a distinct group from the isolates from Germany, Turkey and Syria or previously described *nodC* gene sequences from different geographical regions. Although Bangladeshi isolates formed a strongly separate group in *nodC* gene, it was close to previously described *nodD* type IV (Tian *et al.*, 2010) from China and the Middle East in *nodD* gene.

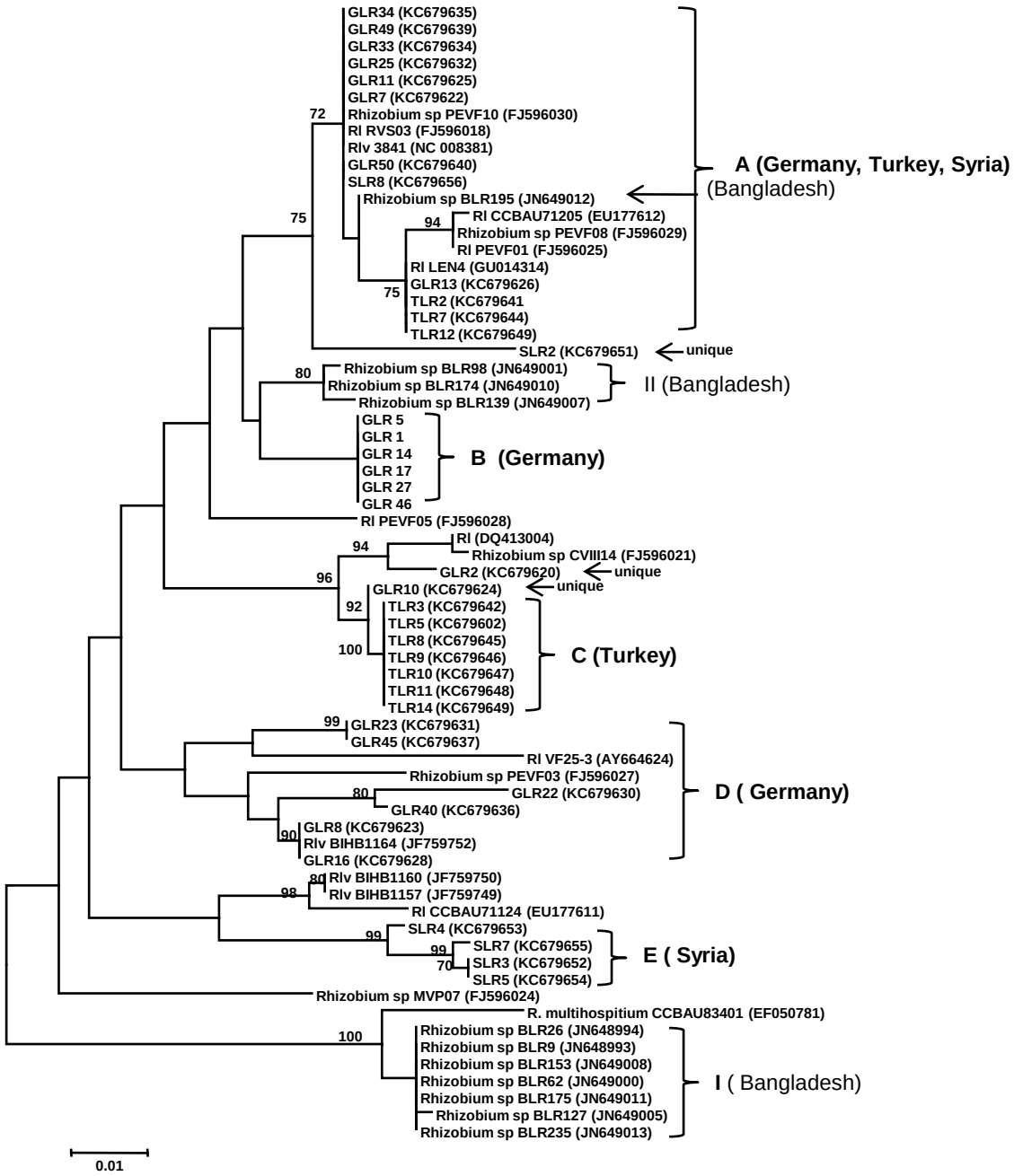


Fig. 27. ML tree from *nodC* gene partial sequences. Bootstrap values indicated when ≥ 70% (1000 replicates). Abbreviation: GLR = German lentil rhizobia, TLR = Turkish lentil rhizobia, SLR = Syrian lentil rhizobia, BLR = Bangladeshi lentil rhizobia, Rlv = *R. leguminosarum* symbiovar *viciae*, Rl = *R. leguminosarum*, A – E = nodulation gene group from German, Turkish and Syrian strains, I – II = nodulation gene group from previous study.

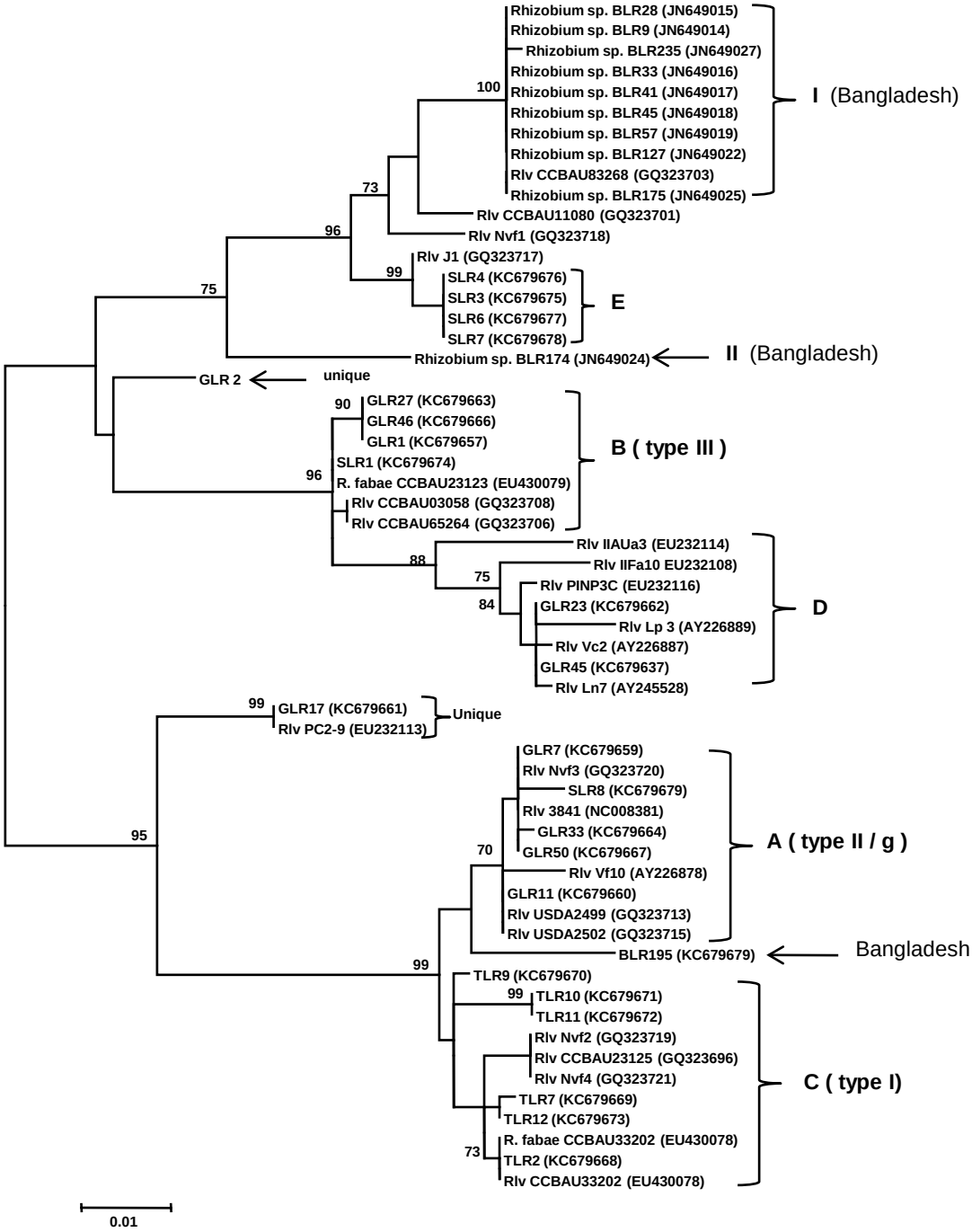


Fig. 28. ML tree from *nodD* gene partial sequences. Bootstrap values indicated when ≥ 70% (1000 replicates). Abbreviations: GLR = German lentil rhizobia, TLR = Turkish lentil rhizobia, SLR = Syrian lentil rhizobia, BLR = Bangladeshi lentil rhizobia, Rlv =*R. leguminosarum* symbiovar *viciae*, Rl = *R. leguminosarum*, A – E = nodulation gene group from present study, I – II = nodulation group from previous study.

3.2.5 Discussion

Rhizobial populations in collected soil and their symbiotic effectivity

Low to moderate lentil-nodulating rhizobial population densities (MPN / g of soil) were found in different localities of Germany (table 3). Although there is no detailed information on the cultivation of lentil in Germany, in 2009 about 0.7% (83,000 ha) of the cultivable land was used for the cultivation of other legumes like pea (http://www.aid.de). In addition to their saprophytic nature (Hirsch, 1996; and reference therein), cultivation of alternative hosts like pea can certainly assist the survival of lentil rhizobia in these soils. This fact may explain our findings regarding moderate density of lentil-nodulating rhizobia in German soil (10^2 – 10^3, cells / g of soil, approximately). Moreover, the variations in terms of rhizobial density found across German localities may be due to differences in soil fertility and land use strategies employed in these collection areas (Mwangi *et al.*, 2011). In nodulation tests, all isolates produced pink nodules with lentil, indicating that they are effective symbionts of this crop. However, in cross inoculation tests all selected isolates formed nodules with others members of the legume tribe *Vicieae*.

Genetic diversity of chromosomal genes and species delineation of lentil rhizobia

We sequenced three protein-coding housekeeping genes (*recA*, *atpD* and *glnII*) that contain valuable information for determining biogeographic processes (Palys *et al.*, 1997; Lan & Reeves, 2001; Vinuesa *et al.*, 2005) and identified six sub-lineages (IVa – IVf) among the 58 isolates from three countries (Turkey, Syria and Germany). All sub-lineages belong to *R. leguminosarum* symbiovr *viciae* and two sub-lineages (IVb and IVd) did not show any great similarity to the sub-lineages that were described earlier within *R. leguminosarum* symbiovr *viciae* (Tian *et al.*, 2012), evidencing that they are new sub-lineages within *R. leguminosarum* symbiovr *viciae*. We found phylogenetic incongruence in reconstructed trees from chromosomal genes. Consequently, result from different analyses (S-H test, recombination analyses, estimation of TMRCA) showed substantial level of recombination among sub-lineages (Fitch, 1997; Shimodaira & Hasegawa, 1999; Bryant & Moulton, 2004; Tian *et al.*, 2010).

Network analyses allow conflicting or alternative phylogenetic histories to account for ambiguities caused by recombination, hybridization, gene conversion and gene transfer (Fitch, 1997). From this analysis we obtained a clear reticulate structure among different sub-lineages from three different countries, but a long edge between Bangladeshi isolates, and isolates from other three countries. In other words, network analyses showed a clear difference between lineage IV and the rest by a long edge, suggesting that lineage IV belongs to separate species (Aguilar *et al.*, 2004; Bailly *et al.*, 2006). Lentil-nodulating rhizobia from Bangladesh were well-separated by long edges from *R. leguminosarum* (those nodulating lentils in different countries of the Mediterranean region) and always formed three distinct lineages without any incongruence in phylogenetic analyses; evidencing that Bangladeshi isolates belong to separate species. Based on protein coding genes (*recA* and *glnII*), lineage IV is genetically similar to the *R. leguminosarum* type strain (> 94%) and form a reticulate structure with *R. leguminosarum* in network analysis suggesting that these sub-lineages belong to this species (Bailly *et al.*, 2006; Valverde *et al.*, 2006; Santillana *et al.*, 2008).

Recombination history in lentil-nodulating rhizobia

The interpretation of population structure for a particular species of bacteria collected from diverse ecological niches heavily depends on the proper interpretation of different evolutionary processes acting on chromosomal and plasmid-borne genes (Lan & Reeves, 2001; Vinuesa *et al.*, 2005). Hence, we used phylogenetic and population genetic approaches to explore the the evolutionary history of lentil rhizobia.

In this study, we found phylogenetic incongruence in reconstructed trees from chromosomal genes, and their topology showed significant differences in S-H tests. Using concatenated sequences, a clear reticulate structure was found in network analyses, coupled with high values of r/m statistics (r/m = 1.52) in recombination analyses. Supporting this pattern, a high number of recombination events (43) was found in three loci. Moreover, relative rates of recombination (ρ/θ) were high (0.269). Furthermore, the estimation of TMRCA with and without recombination differs significantly and reveals different consensus tree topologies. These results indicate the presence of recombination (Fitch, 1997; Shimodaira & Hasegawa, 1999;

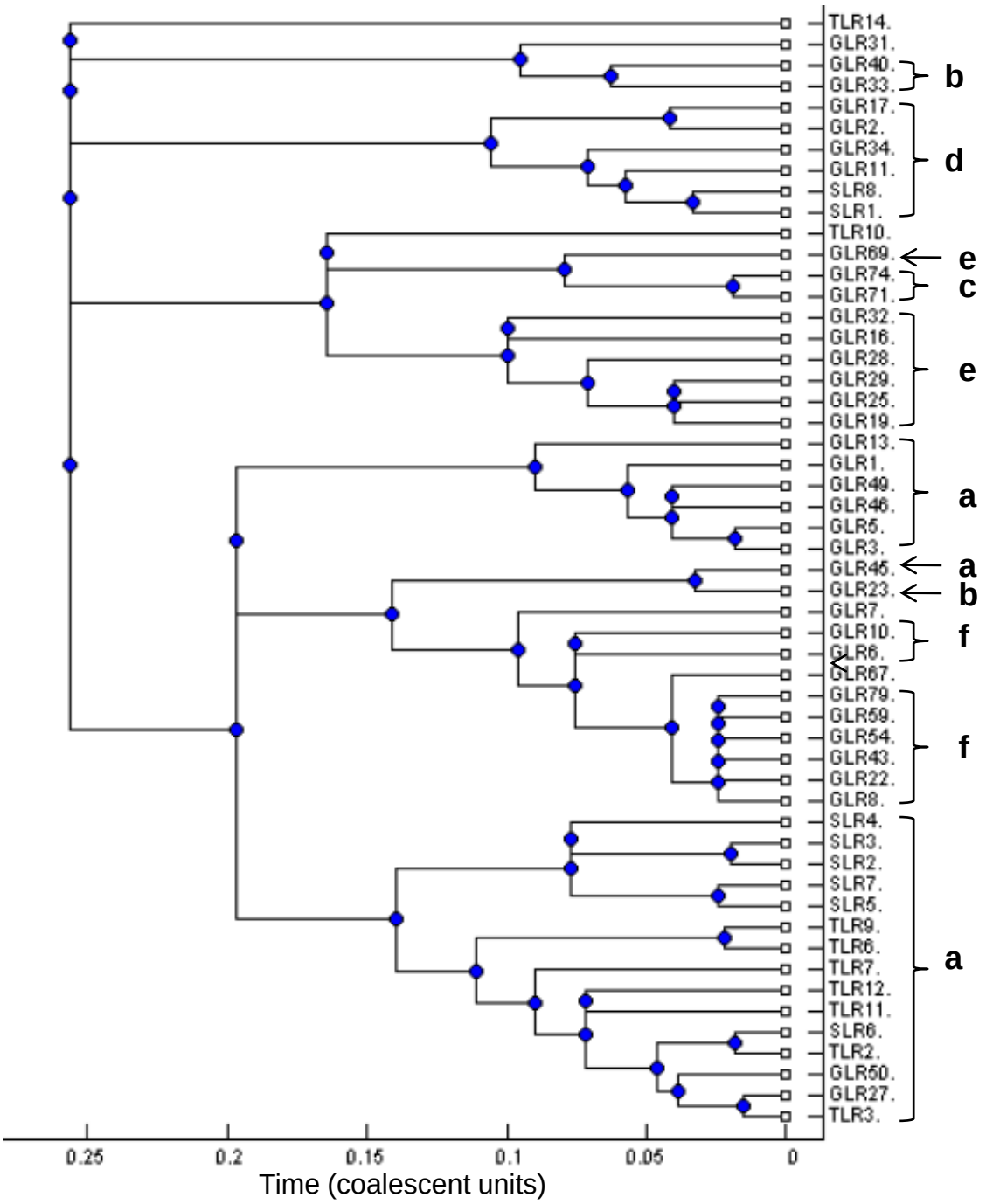


Fig. 29a. Dendogram, a majority rule consensus tree (50%) inferred from concatenated partial sequence of *recA-atpD-glnII* genes using CLONALFRAME allowing recombination. The scale indicate the time in coalescent units. Abbreviations: GLR = German lentil rhizobia, TLR = Turkish lentil rhizobia, SLR = Syrian lentil rhizobia, a – f = sub-lineages within lineage IV.

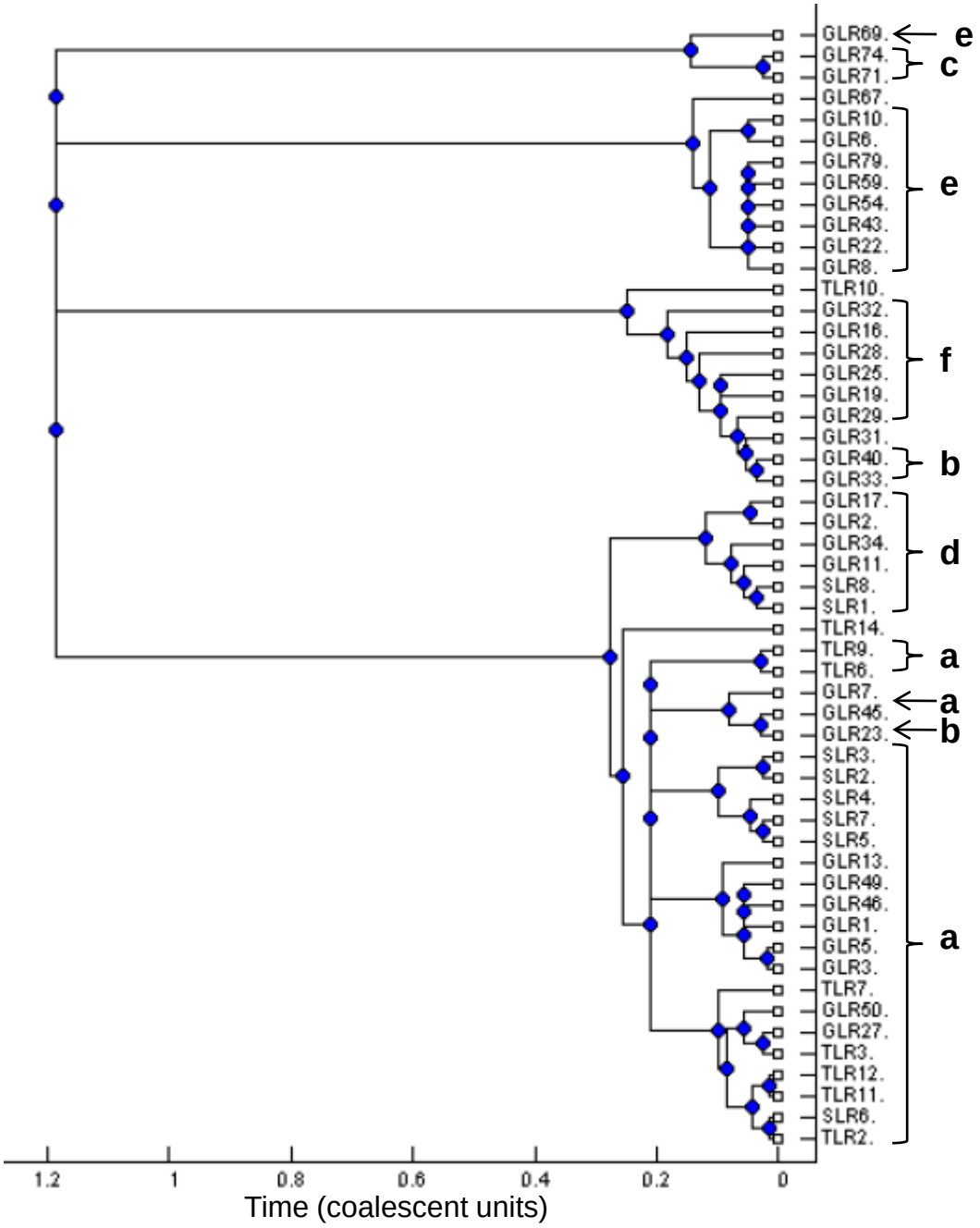


Fig. 29b. Dendogram, a majority role consensus tree (50%), inferred from concatenated partial sequence of *recA-atpD-glnII* genes using CLONALFRAME without recombination. The scale indicates the time in coalescent units. Abbreviations: GLR = German lentil rhizobia, TLR = Turkish lentil rhizobia, SLR = Syrian lentil rhizobia, a – f = sub-lineages within lineage IV.

Bryant & Moulton, 2004; Tian *et al.*, 2010) among the different sub-lineages. The influence of recombination might explain why we did not recover same sub-lineages with equal resolution from phylogenetic and network analyses (Holmes *et al.*, 1999; Didelot & Falush, 2007; den Bakker *et al.*, 2008). Moreover, in the presence of recombination, phylogenetic approaches used for bacterial taxa and species delineation may not shown correct interpretations (Rosselló-Mora & Amann, 2001). Alternatively, by allowing conflicting or alternative phylogenetic histories, network analyses represent valuable tools for resolving ambiguities caused by recombination, hybridization, gene conversion and gene transfer (Fitch, 1997).

Successful horizontal gene transfer in bacteria is assume to provide a s elective advantage to either the host or the gene itself (Gogarten & Townsend, 2005). Therefore, it could be assumed that to cope with diverse conditions, *R. leguminosarum* from the Europe and the Middle East may allow recombination at loci. In contrast, we did not see any phylogenetic incongruence in single gene and concatenated sequence analyses in Bangladeshi isolates under present study and in our previous study (Rashid *et al.*, 2012) which may be due to soil conditions or crop cultivations histories.

Geographic structure of lentil-nodulating rhizobia

Both gene flow and founder effect act differentially at small or large scales to shape the genetic structure of bacteria (Strain *et al.*, 1995; Hagen & Hamrick, 1996; Silva *et al.*, 2005; Vinuesa *et al.*, 2005). The estimation of these parameters in the three different countries reveals high levels of migration within the center of origin of lentil (Turkey-Syria), but low gene flow between this center of origin and Europe (Germany). This result is in contrast to previous study (Silva *et al.*, 2005) where large distances (more than 10,000 km) represented an important barrier for gene flow, while populations separated by 1,000 – 2,000 km did not show geographical sub-division. About 3,500 and 4,000 km separate Germany from Turkey and from Syria, respectively. These distances represent almost one third of the distance proposed by Silva *et al.* (2005).

Direct dispersal (inoculation of legumes) or indirect dispersal (with seed) of rhizobia may occur in legumes (Perez-Ramírez *et al.*, 1998). Moreover, geographical distance and environmental conditions may have a minor effect on bacterial diversity when high levels of recombination are present (Fraser *et al.*, 2009), which may be t he case for lentil-nodulating rhizobia.

Origin and distribution of new lineages in Bangladesh are influenced by symbiosis with lentil grex *pilosae*

For cultivated lentils, Barulina (1930) proposed six geographical groups, or greges, *viz. europeae*, *asiaticae*, *intermediae*, *subspontanea*, *aethiopicae* and *pilosae* (Cubero, 1981; 2009; and references therein). Interestingly, among the six greges of cultivated lentil, three groups are restricted to very specific areas. For instance, *pilosae* is endemic to the Indian sub-continent, *aethiopicae* to Ethiopia and Yemen, and *subspontanea* to Afghanistan. This study and others on lentil rhizobia show clearly that *R. leguminosarum* is the main symbiont of all lentil greges except *pilosae* in the Indian sub-continent.

Although Eastern Turkey and Northern Syria are the area of lentil domestication (Ladizinsky, 1979; Lev-Yadun *et al.*, 2000; Zohary & Hopf, 2000; Cubero *et al.*, 2009), Barulina (1930) stated that the Himalaya-Hindu Kush corresponds to the centre of origin for small seeded *microsperma* lentils because of the presence of a higher proportion of endemic varieties. *Pilosae* has a strong pubescence which is absent in other lentils (Cubero *et al.*, 2009) and a flowering asynchrony, and has not overlapped with any other greges of lentil during the history of domestication and cultivation (Barulina, 1930). This grex may also have specific genetic characteristics like nod factor receptors that allow for a successful symbiosis with the new *Rhizobium* species or lineages found in Bangladesh. Thus, *pilosae* and their symbionts may have coevolved in the Indian sub-continent (Perret *et al.*, 2000; Aguilar *et al.*, 2004). It is therefore possible that we found new lineages (Rashid *et al*., 2012) of rhizobia from Bangladesh due to the significant influence of the *pilosae* grex on their symbiotic partners.

Rhizobium leguminosarum is the original symbiont of lentils

In agreements with other studies (Moawad & Beck, 1991; 1998; Laguerre *et al.*, 1992; Geniaux & Amargr, 1993; Tegegn, 2006), we found *R. leguminosarum* in the centre of origin of lentil and in countries where lentil had been introduced (e.g. Germany). We not only found the same *R. leguminosarum* species but also the same chromosomal genotype (e.g. sub-lineage IVa) in three different countries.

Hence, it could be assumed that *R. leguminosarum* is the original symbiont of lentils. The dispersal of rhizobia with legume seeds is a well-accepted hypothesis (Perez-Ramírez *et al.*, 1998; Aguilar *et al.*, 2004; Álvarez-Martinez *et al.*, 2009). This mode is considered to be the most important among the indirect ways of rhizobium dispersal

(Hirsch 1996; and references therein). Lentil is the oldest crop and has remained popular from ancient times until to date, and found all over the world. Thus it could be assumed that lentils seed might played a significant role for initial dispersion of *R. leguminosarum* symbiovar *viciae* in different countries.

Phylogenetic incongruence between chromosomal and nodulation genes

Rhizobial species diversity should be described not only based on genetic markers located on chromosomes but also on plasmid-borne nodulation genes (Graham *et al.*, 1991; Amarger *et al.*, 1997; Wang *et al.*, 1999; Laguerre *et al.*, 2001; Silva *et al.*, 2005). The *nodC* and *nodD* genes determine host range, host promiscuity and the relationships between host plants and rhizobia (Laguerre *et al.*, 2001; Zeze *et al.*, 2001; Iglesias *et al.*, 2007). In our study, phylogenetic analyses based on nodulation genes confirm that these isolates belong to symbiovar *viciae* and they have *nodC* and *nodD* genotypes which have previously been described from Europe, Middle East and China. We also detected new groups within this symbiovar.

There was no congruence between chromosomal and nodulation genotypes, which may be due to frequent lateral transfer of nodulation genes between different rhizobial chromosomal genotypes (Sprent, 1994; Young & Haukka, 1996). However, phylogenetic analyses based on nodulation genes (*nodC* and *nodD*) revealed similar tree topologies correlated mostly with ecological regions. Probably as a consequence of both host and soil microhabitats (Sprent, 1994), we detected new nodulation genotypes (or groups) that were supported by both nodulation genes. These results support the hypothesis that plasmid-borne characters in bacteria change rapidly according to their adaptation to particular environments. Nonetheless, nodulation genes from Europe and the Middle East did not show great similarities to the isolates from Bangladesh, suggesting that the latter have an independent origin on the Indian subcontinent. A similar hypothesis has been proposed for the origin of *R. etli* (Aguilar *et al.*, 2004). Out of the three distinct lineages, one lineage (lineage I) is found all over Bangladesh (Rashid *et al.*, 2012), suggesting that it may have been distributed with *pilosae* seeds.

Conclusions

By analyzing lentil-nodulating rhizobia from four countries in two different continents we found four different lineages of rhizobia, of which three are new. These three new lineages of rhizobia are endemic to Bangladesh and have been coevolved with grex *pilosae* of lentil. The presence of common genotypes of *R. leguminosarum* with lentil in different countries, suggest that *R. leguminosarum* is the original symbiont of lentil. Further research is needed in order to further examine the genetic diversity and population structure of lentil-nodulating rhizobia in different geographical regions.

4. Conclusions and general discussion

In this study lentil-nodulating rhizobia were isolated from countries at the centre of the origin of lentil (Turkey and Syria) and from countries where lentil was introduced later (Germany and Bangladesh). The collected rhizobial isolates were analyzed based on multilocus sequence analyses, DNA fingerprinting, and phenotypic characteristics. Phylogenetic and population genetic approaches have been used to interpret the data in terms of rhizobial diversity and taxonomic status of rhizobia associated with lentil. To our knowledge this is the first detailed study of lentil-nodulating rhizobia. This study found four different lineages, three of which are novel. Moreover, the study also found that evolutionary forces such as recombination played a greater role in the diversification of *R. leguminosarum* than mutation. The results presented in this dissertation highlight two major findings: the presence of three new endemic lineages in Bangladesh (Indian sub-continent) that are close to *R. etli* but distinct from the *R. leguminosarum-R etli* complex, and the origin and distribution of a cosmopolitan lineage of lentil symbiont (*R. leguminosarum*) to different countries.

4.1 Genetic diversity of rhizobia nodulating lentil in Bangladesh

Multilocus sequence analysis (MLSA) is considered to be an effective approach for describing bacterial species diversity and speciation events. It is also a key approach for describing closely-related bacterial species (Konstantinidis *et al.*, 2006; and references therein; Martens *et al.*, 2008; and references therein). Besides their functional importance, some housekeeping genes like *recA, atpD* and *glnII* also provide useful information for describing closely-related bacterial species that cannot be resolved using 16S rRNA gene sequences (Bromofeld *et al.*, 2010; and references therein). Using a MLSA approach, our study revealed that three distinct lineages are associated with lentil nodulation in Bangladesh. Lineages I and II were distributed throughout Bangladesh, whereas lineage III was confined to the South East part of the country. Three protein-coding genes from the isolates taken from Bangladesh showed similarity values of 89.1 – 92.8% and 92.1 – 94.8% to *R. leguminosarum* and *R.etli*, respectively. Moreover, the genetic similarity values between these new lineages were 91 – 92% for *recA*, 92 – 94% for *glnII* and 89 – 94% for *atpD*. The genetic similarity levels between the new lineages are comparable to those found between *R. etli* and *R. leguminosarum*, suggesting species status for these lineages from Bangladesh. These lineages are close to *R. etli* but significantly different from this species based on DNA

sequence information from the four housekeeping genes. This conclusion is also highly supported by DNA fingerprinting and phenotypic data. Different phylogenetic analyses like neighbor-joining, maximum likelihood and Bayesian inference methods showed similar tree topologies. Therefore, from this study it is clear that at least three different rhizobial lineages are involved in lentil nodulation in Bangladesh that belong to new species within the genus *Rhizobium*.

4.2 Origin of rhizobia nodulating lentil in Bangladesh

The host has a significant effect on its symbiont, and the dispersion of legume-nodulating bacteria may have been facilitated by the export of plants and seeds (Perez-Ramírez *et al.*, 1998; Aguilar *et al.*, 2004). It is assumed that *R. etli* originated in the Americas with its host *Phaseolus vulgaris* and was distributed with its seed to other countries around the world (Aguilar *et al.*, 2004). Although the lineages from Bangladesh were close to *R. etli,* its host *Phaseolus vulgaris* is neither widely known nor cultivated in Bangladesh, suggesting a low probability for the distribution of these new lineages to Bangladesh with *Phaseolus vulgaris* seeds. However, Bangladesh has a long history for lentil cultivation. Among the six cultivated greges of lentil, *pilosae* is the only cultivated grex widely available throughout Bangladesh and other parts of the Indian sub-continent. We obtained new lineages of rhizobia from endemic *pilosae* lentils that were absent from other parts of the world. Parallel and independent evolution may occur in different locations in strains of pathogenic and symbiotic bacteria, (Schuster & Coyne, 1974; Geffroy *et al.*, 1999; Aguilar *et al.*, 2004; Wolde-Meskel *et al.*, 2005) suggesting that these new lineages may potentially be endemic to Bangladeshi soil and may have coevolved with locally grown *pilosae* lentils.

4.3 Genetic diversity of lentil-nodulating rhizobia from the Middle East and Germany are greatly influenced by recombination

Although the 16S rRNA gene sequences from the 58 isolates from Germany, Turkey and Syria were 99 – 100% similar to *R. leguminosarum*, we identified six sub-lineages (Iva – IVf) by analyzing three housekeeping genes. Out of these, two sub-lineages (IVb and IVd) showed substantial differences to other sub-lineages and strains of *R. leguminosarum* that had been described earlier (Tian *et al.*, 2010). Hence, they might represent new sub-lineages within *R. leguminosarum*. However, all sub-lineages were not equally well-resolved in phylogenetic and population analyses. Therefore, our study

further supports the view that the interface between populations and species should be explored using both population genetics and phylogenetic approaches (Lan & Reeves, 2001; Vinuesa *et al.*, 2005).

Different analyses (recombination, mutation, tree topologies, structure) of three protein-coding genes showed phylogenetic incongruence among the reconstructed trees and significant differences in their topologies, and a substantial influence of recombination (Fitch, 1997; Shimodaira & Hasegawa, 1999; Bryant & Moulton, 2004; Tian *et al.*, 2010) in different sub-lineages. We identified a maximum of six sub-lineages in ML analyses that were not found in other analyses; this might be due to a violation of the phylogenetic assumption of bifurcation by the presence of recombination (Holmes *et al.*, 1999; Didelot & Falush, 2007; den Bakker *et al.*, 2008) in the different sub-lineages. By allowing conflicting or alternative phylogenetic histories, like recombination, network analyses represent valuable tools for resolving ambiguities (Fitch, 1997; Vinuesa *et al.*, 2005; and references therein). This analysis produced a reticulate structure among sub-lineages; suggesting a great influence of recombination on *R. leguminosarum*. In contrast, recombination was apparently absent in three new lineages found in Bangladesh, and we did not see any phylogenetic incongruence and reticulate structure in single-gene and concatenated sequence analyses. This might be due to the influence of soil conditions or crop cultivation histories in Bangladesh.

4.4 Species delineation of lentil rhizobia

The ML analysis and genetic similarity showed that lentil rhizobia from Bangladesh belong to three new lineages, and found a cosmopolitan lineage in the Middle East and central Europe. Similar result were also observed in network analysis, which showed a long edge between the German-Turkish-Syrian isolatesand Bangladeshi isolates and a reticulate structure among six sub-lineages within *R. leguminosarum*, suggesting that the Bangladeshi isolates belong to separate species (Aguilar *et al.*, 2004; Bailly *et al.*, 2006). Although ML analysis found six sub-lineages in the Middle Eastern and German isolates, they form a reticulate structure with *R. leguminosarum*; evidencing that all sub-lineages belong to this species.

4.5 *Rhizobium leguminosarum* is the original symbiont of lentils

Most studies on the origin and distribution of cultivated lentil indicated that the center of origin of lentil is located in the Near East and Central Asia (Shandu & Singh, 2007; and

references therein). We found *R. leguminosarum* in the center of origin (Syria and Turkey) and in a country where lentil had been introduced later (Germany). Previous studies also found *R. leguminosarum* in the Mediterranean region, European countries, different African countries and in the Americas (Moawad & Beck, 1991; Laguerre *et al.*, 1992; Geniaux & Amargr, 1993; Moawad *et al.*, 1998; Tegegn, 2006). Moreover, both *macrosperma* and *microsperma* lentils, with the exception of the grex *pilosae*, are present in Turkey (www.economy.gov.tr), but we did not find any rhizobial strains that are similar to the ones isolated from Bangladesh. In contrast, we not only found *R. leguminosarum* in the center of origin but also in central Europe, including common chromosomal genotypes such as sub-lineage IVa. Hence, it could be assumed that *R. leguminosarum* is the original symbiont of lentil.

Legume seed is a major carrier of rhizobia (Perez-Ramírez *et al.*, 1998; Aguilar *et al.*, 2004), and lentils seed therefore might also carry rhizobia on their testa. Lentil is one of the oldest cultivated crops and has remained popular until now. Due to its popularity and ancient origin, it can be assumed that after its domestication in the cradle of agriculture lentil symbionts quickly dispersed to Europe and Africa with its host (Zohary & Hopf, 1993; Shandu & Singh, 2007; Sonnante *et al.*, 2009). The presence of common chromosomal genotypes in both the center of lentil origin and in countries to which lentil was introduced later suggest a common phylogenetic origin of *R. leguminosarum.* Therefore, *the* spread of lentils may have helped the initial dispersion of *R. leguminosarum* from the cradle of agriculture to other countries.

4.6 Genetic diversity of nodulation genes

Nodulation factors (Nod) help to establish legume-rhizobium symbiosis (Laguerre *et al.*, 2001), and Nod factor synthesis depends on different nodulation genes. Nodulation genes like *nodA, nodC* and *nodD* are widely used as symbiotic marker for describing rhizobial diversity (Perret *et al.*, 2000; Laguerre *et al.*, 2001; Iglesias *et al.*, 2007; Tian *et al.*, 2010; and references therein). Although an analysis of chromosomal genes found three distinct lineages in Bangladesh, phylogenetic analyses of three nodulation genes (*nodA, nodC and nodD*) detected two genotypes / groups among these three lineages. Nodulation gene sequences showed high similarities to those of *R. leguminosarum* symbiovar *viciae*, but they still formed a separate genotypes supported by high bootstrap values (79 – 99%). Phylogenetic analyses of nodulation genes from the Middle Eastern and the central European isolates also support a similarity between the

isolates and *R. leguminosarum* symbiovar *viciae.* We also detected new groups within this symbiovar that were biased to their geographical origin. There was no congruence between nodulation gene and chromosomal gene phylogeny, which may be due to horizontal / lateral nodulation gene transfer between different chromosomal lineages and sub-lineages (Ochman *et al.*, 2000; Laguerre *et al.*, 2001; Han *et al.*, 2010; Degefu *et al.*, 2011). In most of the cases, the phylogeny of nodulation genes showed a positive correlation with their ecological origin, which may be a consequence of both the host and the soil microhabitats (Sprent, 1994).

Like the lentil-nodulating rhizobia from different countries (Bangladesh, Germany, Turkey and Syria), *R. leguminosarum* symbiovar *viciae* 3841 also nodulates lentil effectively and produces effective nodules with lentil in our nodulation test. Therefore, we can predict that the symbiotic performance of lentil isolates lies within the variability of *R. leguminosarum* symbiovar *viciae*.

4.7 Specific conclusion

Our present study found four different lineages of rhizobia associated with lentil nodulation in different countries, of which three are new. These three new lineages are closely related to *R. etli* but phylogentically distinct enough to warrant new species status within the genus *Rhizobium*. Phylogenetic conclusions are highly supported by DNA fingerprinting data and phenotypic data. Therefore, our study found three species-level new lineages in Bangladesh. The *pilosae* lentil is endemic to Bangladesh and the Indian sub-continent, and we isolated new lineages of rhizobia from *pilosae* root nodules. Both the new rhizobial lineages and the lentil grex are absent from other parts of the world, suggesting that these new rhizobial lineages are endemic to Bangladesh or the Indian sub-continent and co-evolved with *pilosae* lentil. This study showed a common *R. leguminosarum* genotype present in Turkey and Syria, which are at the origin of all *macrosperma* and *microsperma* lentil greges with the exception of *pilosae,* and in central Europe (Germany), suggesting that *R. leguminosarum* is the original symbiont of lentil. The presence of a common lineage and sub-lineage at centre of origin and in other countries suggested that lentil seed might have played a major role in the initial distribution of this symbiont to different countries from the Middle East on to other countries. Further research with more isolates from more geographical locations and analyses of genetic markers from different lentil greges are needed to examine the genetic diversity and population structure of lentil-nodulating rhizobia.

5. References

Adekambi T & Drancourt M (2004) Dissection of phylogenetic relationships among rapidly growing *Mycobacterium* species by 16S rRNA, *hsp65*, *sodA*, *recA* and *fpoB* gene sequencing. *Int J Syst Evol Micr* **54**: 2095 – 2105.

Aguilar OM, Riva O & Peltzer E (2004) Analysis of *Rhizobium etli* and of its symbiosis with wild *Phaseolus vulgaris* supports coevolution in centers of host diversification. *P Natl Acad Sci USA* **101**: 13548 – 13553.

Álvarez-Martinez ER, Valverde A, Ramírez-Bahena MH, García-Fraile P, Tejedor C, Mateos PF, Santillana N, Zúñiga D, Peix A & Velázquez E (2009) The analysis of core and symbiotic genes of rhizobia nodulating *Vicia* from different continents reveals their common phylogenetic origin and s uggests the distribution of *Rhizobium leguminosarum* strains together with *Vicia* seeds. *Arch Microbiol* **191**: 659 – 668.

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

An DS, Im WT, Yang HC & Lee ST (2006) *Shinella granuli* gen. nov., sp. nov., and proposal of the reclassification of *Zoogloea ramigera* ATCC 19623 as *Shinella zoogloeoides* sp. nov. *Int J Syst Evol Micr* **56**: 443 – 448.

Anyango B, Wilson KJ, Beynon JL & Giller KE (1995) Diversity of rhizobia nodulating *Phaseolus vulgaris* in two Kenyan soils of contrasting pHs. *Appl Environ Microb* **61**: 4016 – 4021.

Bailly X, Olivieri I, de Mitas S, Cleyet-Marel J & Béna G (2006) Recombination and selection shape the molecular diversity pattern of nitrogen-fixing *Sinorhizobium* sp. associated to *Medicago*. *Mol Ecol* **15**: 2719 – 2734.

Baldani JI, Baldani VLD, Seldin L & Döbereiner J (1986) Characterization of *Herbaspirillum seropedicae* gen. nov., sp. nov., a r oot-associated nitrogen-fixing bacterium. *Int J Syst Bacteriol* **36**: 86 – 93.

Baldwin IL & Fred EB (1929) Root-nodulating bacteria of leguminosae. *J Bacteriol* **17**: 141 – 150.

Barton NH, Briggs DEG, Eisen JA, Goldstein DB & Patel NH (2007) Evolution. (eds.), Cold spring harbor laboratory press, Europe.

Barulina H (1930) *Lentil of the U.S.S.R. and of other countries.* Vsesoíúžnaíâ`akademiíâ` sel'sko khozíaî`stvennyk, Lenina, Russia.

Bergey DH, Harrison FC, Breed RS, Hammer BW & Huntoon FM (1923) *Bergey's Manual of Determinative Bacteriology.* The Williams & Wilkins, Baltimore.

Beringer J (1974) R factor transfer in *Rhizobium leguminosarum*. *J Gen Microbiol* **84**: 88 – 198.

Black M, Moolhuijzen P, Chapman B, Barrero R, Howieson J, Hungria M & Bellgard M (2012) The Genetics of symbiotic nitrogen fixation: comparative genomics of 14 rhizobia strains by resolution of protein clusters. *Genes* **3**: 138 – 166.

Brenner DJ, Staley JT & Krieg NR (2005) Classification of prokaryotic organism and concept of bacterial speciation. *Bergey's manual of systematic bacteriology* Vol. 2 (Brenner DJ, Krieg NR & Staley JT, eds), pp. 27 – 32. Springer, New York.

Brockwell J (1963) Accuracy of a plant-Infection technique for counting populations of *Rhizobium trifolii*. *Appl Microbiol* **11**: 377 – 383.

Bromofeld ESP, Tambong JT, Cloutier S, Prevost D, Laguerre G, van Berkum P, Thi PV, Assabgui R & B arran LR (2010) *Ensifer*, *Phyllobacterium* and *Rhizobium* species occupy nodules of *Medicago sativa* (alfalfa) and *Medicago alba* (sweet clover) grown at a Canadian site without a history of cultivation. *Microbiology* **156**: 505 – 520.

Brown JR & Doolittle WF (1995) Root of the universal tree of life based on ancient amino-acyl-tRNA synthesize gene duplications. *P Natl Acad Sci USA* **92**: 2441 – 2445.

Brown JM & ElDabaje R (2009) PuMA: Bayesian analysis of partitioned (and unpartitioned) model adequacy. *Bioinformatics* **25**: 537 – 538.

Bruneau A, Mercure M, Lewis GP & Herendeen PS (2008) Phylogenetic patterns and diversification in the *Caesalpinioid* legumes. *Botany* **86**: 697 – 718.

Bruno WJ, Socci ND & Halpern AL (2000) Weighted neighbor joining: a l ikelihood-based approach to distance-based phylogeny reconstruction. *Mol Biol Evol* **17**: 189 – 197.

Bryant D & Moulton V (2004) Neighbor-net: an ag glomerative method for the construction of phylogenetic networks. *Mol Biol Evol* **21**: 255 – 265.

Chahota RK, Kishore N, Dhiman KC, Sharma TR & Sharma SK (2007) Predicting transgressive segregants in early generation using single seed descent method-derived micromacrosperma genepool of lentil (*Lens culinaris* Medikus). *Euphytica* **156**: 305-310.

Chen WP & Kuo TT (1993) A simple and rapid method for the preparation of gram negative bacterial genomic DNA. *Nucleic Acids Res* **21**: 2260.

Chen WX, Yan GH & Li JL (1988) Numerical taxonomic study of fast-growing soybean rhizobia and a proposal that *Rhizobium fredii* be assigned to *Sinorhizobium* gen. nov. *Int J Syst Evol Micr* **38**: 392 – 397.

Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG & Thompson JD (2003) Multiple sequence alignment with the Clustal series of programs *Nucl Acids Res* **31**: 3497 – 3500.

Chun J & Hong SG (2010) Methods and programs for calculation of phylogenetic relationships from molecular sequences. *Molecular Phylogeny of Microorganisms* (Oren A & Papke RT, eds.), Caister Academic Press, Norfolk, UK.

Conn HJ (1942) Validity of the genus *Alcaligenes*. *J Bacteriol* **44**: 353 – 360.

Cubero JI (1981) Taxonomy, distribution and evolution of the lentil and its wild relatives. *Lentils* (Witcombe J & Erskine W, eds), pp. 15 – 38. Nijhoff and Junk Publishers, Boston.

Cubero JI, de la Vega MP & Fratini R (2009) Origin, taxonomy and domestication. *The lentil: Botany, production and uses* (Erskine W, Muehlbauer F, Sarker A & Sharma B, eds.), pp. 13 – 33. CAB International, Oxfordshire.

Cullimore JV, Ranjeva R & Bono JJ (2001) Perception of lipo-chitooligosaccharidic Nod factors in legumes. *Trends Plant Sci* **6**: 24 – 30

Cummings SP, Gyaneshwar P, Vinuesa P *et al.,* (2009) Nodulation of *Sesbania* species by *Rhizobium* (*Agrobacterium*) strain IRBG74 and other rhizobia. *Environ Microbiol* **11**: 2510 – 2525.

De Bruijn FJ (1992) Use of repetitive (repetitive extragenic palindromic and enterobacterial repetitive intergeneric consensus) sequences and the polymerase chain reaction to fingerprint the genomes of *Rhizobium meliloti* isolates and other soil bacteria. *Appl Environ Microb* **58**: 2180 – 2187.

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

Degefu T, Wolde-Meskel E & Frostegård A (2011) Multilocus sequence analyses reveal several unnamed *Mesorhizobium* genospecies nodulating *Acacia* species and *Sesbania sesban* trees in Southern regions of Ethiopia. *Syst Appl Microbiol* **34**: 216 – 226.

den Bakker HC, Didelot X, Fortes ED, Nightingale KK & Wiedmann M (2008) Lineage specific recombination rates and microevolution in *Listeria monocytogenes*. *BMC Evol Biol* **8**: 277 – 289.

Dhingra KK, Sekhon HS, Sandhu PS & Bhandari SC (1988) Phosphorus-Rhizobium interaction studies on biological nitrogen fixation and yield of lentil. *J Agr Sci* **110**: 141 – 144.

Didelot X & Falush D (2007) Inference of bacterial microevolution using multilocus sequence data. *Genetics* **175**: 1251 – 1266.

Doyle JJ (1998) Phylogenetic perspectives on no dulation: an evolving view of plants and symbiotic bacteria. *Trends Plant Sci* **3**: 473 – 478.

Doyle JJ (2011) Phylogenetic perspectives on t he origins of nodulation. *Mol Plant Microbe Interact* **24**: 1289 – 1295.

Doyle JJ & Luckow MA (2003) The rest of the iceberg: legume diversity and evolution in a phylogenetic context. *Plant Physiology* **131**: 900 – 910.

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

Eardly BD, Nour SM, van Berkum P & Selander RK (2005) Rhizobial 16S rRNA and *dnaK* mosaicism and the uncertain phylogenetic placement of *Rhizobium galegae*. *Appl Environ Microb* **71**: 1328 – 1335.

Erskine W (1997) Lessons for breeders from land races of lentil. *Euphytica* **93**: 107 – 112.

Erskine W, Smartt J & Muehlbauer FJ (1994) Mimicry of lentil and the domestication of common vetch and grass pea. *Econ Bot* **48**: 326 – 332.

Excoffier L, Laval G & Schneider S (2005) Arlequin ver. 3.5: an integrated software package for population genetics data analysis. *Evol Bioinform* **1**: 47 – 50.

Excoffier L, Somouse P & Quattro J (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**: 479 – 491.

Falush D, Stephens M & Pritchard J (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**: 1567 – 1587.

Fåhreus G (1957) The infection of clover root hairs by nodule bacteria studied by a simple glass slide technique *J Gen Microbiol* **16**: 374 – 381.

FAOSTAT-Agriculture (2004) *Food and agricultural commodities production*. Food and agricultural organization. Rome.

FAOSTAT-Agriculture (2007) *Food and agricultural commodities production*. Food and agricultural organization. Rome.

FAOSTAT-Agriculture (2010) *Food and agricultural commodities production*. Food and agriculture organization. Rome.

Fellay R, Perret X, Viprey V, Broughton WJ & Brenner S (1995) Organization of host-inducible transcripts on the symbiotic plasmid of *Rhizobium* sp. NGR234. *Mol Microbiol* **16**: 657 – 667.

Felsenstein J (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol* **17**: 368 – 376.

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

Felsenstein J (1988) Phylogenies and quantitative characters. *Ann Rev Ecol Syst* **19**: 445 – 471.

Fitch W (1997) Networks and viral evolution. *J Mol Evol* **44**: 65 – 75.

Fox GE, Magrum LJ, Balch WE, Wolfe RS & Woese CR (1977) Classification of methanogenic bacteria by 16s ribosomal RNA characterization. *Proc Natl Acad Sci USA* **74**: 4537 – 4541.

Frank B (1889) *Über die Pilzsymbiose der Leguminosen.*Vol.19*,* Verlog von Paul Parey Berlin.

Fraser C, Alm EJ, Polz MF, Spratt BJ & Hanage WP (2009) The bacterial species challenge: making sense of genetic and ecological diversity. *Science* **323**: 741 – 746.

Fred EB, Baldani JI & McCoy E (1932) *Root nodule bacteria and legume plants* University of Wisconsin, USA.

Futuyma DJ (2009) *Evolution.* Sinauer Associates Sunderland, Massachusetts.

Gascuel O (1997) Concerning the NJ algorithm and its unweighted version, UNJ. *Mathematical hierarchies and biology* (Mirkin B, McMorris FR, Roberts FS & Rzhetsky A, eds), pp. 149 – 170. American Mathematical Society, Washington, DC.

Gaunt MW, Turner SL, Rigottier-Gois L, Lloyd-Macgilp SA & Young JPW (2001) Phylogenies of *atpD* and *recA* support the small subunit rRNA-based classification of rhizobia. *Int J Syst Evol Micr* **51**: 2037 – 2048.

Geffroy V, Sicard D, de Oliveira JCF, Sévignac M, Cohen S, Gepts P, Neema C, Langin T & Dron M (1999) Identification of an ancestral resistance gene cluster involved in the coevolution process between *Phaseolus vulgaris* and its fungal pathogen *Colletotrichum lindemuthianum*. *Mol Plant Microbe Interact* **12**: 774 – 784.

Gelman A & Rubin DB (1992) Inference from iterative simulation using multiple sequences. *Stat Sci* **7**: 457 – 511.

Geniaux E & Amarger N (1993) Diversity and stability of plasmid transfer in isolates from a s ingle field population of *Rhizobium leguminosarum* bv. *viciae*. *FEMS Microbiol Ecol* **102**: 251 – 260.

Giraud E, Moulin L, Vallenet D *et al.,* (2007) Legumes symbioses: Absence of Nod genes in photosynthetic bradyrhizobia. *Science* **316**: 1307 – 1312.

Glazunova, Raoult D & Roux V (2009) Partial sequence comparison of the *rpoB*, *sodA*, *groEL* and *gyrB* genes within the genus *Streptococcus. Int J Syst Evol Micr* **59**: 2317 – 2322.

Goodfelow M (1971) Numerical taxonomy of some noncardioform bacteria. *J Gen Microbiol* **69**: 33 – 80

Gogarten JP & Townsend JP (2005) Horizontal gene transfer, genome innovation and evolution. *Nat Rev Microbiol* **3**: 679 – 687.

Gonzalez J & Wink M (2010) Genetic differentiation of the Thorn-tailed Rayadito (*Aphrastura spinicauda*) revealed by ISSR profiles suggest multiple paleorefugia and high recurrent gene flow. *IBIS* **152**: 761 – 774.

Graham PH (1976) Identifcation and classifcation of root nodule bacteria. *Symbiotic Nitrogen Fixation in Plants* (Nutman PS, ed), Cambridge University Press, London.

Graham PH & Parker CA (1964) Diagnostic features in the characterization of the root-nodule bacteria of legumes. *Plant Soil* **20**: 383 – 396.

Graham PH, Sadowsky MJ & Keyser HH (1991) Proposed minimal standards for the description of new genera and species of root- and stem-nodulating bacteria. *Int J Syst Evol Micr* **41**: 582 – 587.

Graur D & Li WH (2000) *Fundamentals of molecular evolution.* Massachusetts, Sinauer Associates, Sunderland.

Guttman DS & Dykhuizen DE (1994) Clonal divergence in *Escherichia coli* as a result of recombination, not mutation *Science* **266**: 1380 – 1383.

Gyaneshwar P, Hirsch, AM, Moulin L. et al., (2011) Legume-nodulating Betaproteobacteria: diversity, host range and future prospects. *Mol Plant Microbe Interact* **24**: 1276 – 1288.

Hagen MJ & Hamrick JL (1996) A hierarchical analysis of population genetic structure in *Rhizobium leguminosarum* bv. *trifolii*. *Mol ecol* **5**: 177 – 186.

Hall BG (2008) *Phylogenetic Trees Made Easy: A how to manual.* Sinauer Associates.

Hall TA (1999) BioEdit: a us er friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* **41**: 95 – 98.

Hamady M & Knight R (2009) Microbial community profiling for human microbiome projects: tools, techniques, and challenges. *Genome Res* **19**: 1141 – 1152.

Han TX, Tian CF, Wang ET & Chen WX (2010) Associations among rhizobial chromosomal background, nod genes and host plants based on the analysis of symbiosis of indigenous rhizobia and w ild legumes native to Xinjiang. *Microbial Ecol* **59**: 311 – 323.

Harrison, CJ & Langdale JA (2006) A step by step guide to phylogeny reconstruction. Plant J **45**: 561 – 572

Haukka K, Lindstrom K & Young JP (1998) Three phylogenetic groups of *nodA* and *nifH* genes in *Sinorhizobium* and *Mesorhizobium* isolates from leguminous trees growing in Africa and Latin America. *Appl Environ Microb* **64**: 419 – 426.

Hellriegel H & Wilfarth H (1888) Untersuchungen über die Stickstoffnahrung der Gramineon und Leguminosen. Beilageheft zu der Ztschr. *Ver. Rübenzucker-Industrie Deutschen Reichs*.

Hennig W (1966); Davis D & Zangerl R (translator) *Phylogenetic Systematics* (Reissue I, ed.), pp. 72 – 77, Board of Trustees of the University of Illinois, USA.

Herrera-Cervera JA, Sanjuan-Pinilla JM, Oliveres J & Sanjuan J (1998) Cloning and identification of conjugative transfer origins in the *Rhizobium meliloti* genome. *J Bacteriol* **180**: 4583 – 4590.

Hirsch PR (1996) Population dynamics of indigenous and genetically modified rhizobia in the field. *New Phytol* **133**: 159 – 171.

Holmes EC, Urwin R & Maiden MC (1999) The influence of recombination on the population structure and evolution of the human pathogen *Neisseria meningitidis*. *Mol Biol Evol* **16**: 741 – 749.

Hou BC, Wang ET, Li Y, Jia RZ, Chen WF, Man CX, Sui XH & Chen WX (2009) Rhizobial resource associated with epidemic legumes in Tibet. *Microbial Ecol* **57**: 69 – 81.

Hudson RR, Slatkin M & Maddison WP (1992) Estimation of levels of gene flow from DNA sequence data. *Genetics* **132**: 583 – 589.

Hynes MF & O'Connell MP (1990) Host plant effect on competition among strains of *Rhizobium leguminosarum*. *Can J Microbiol* **36**: 864 – 869.

Ibañez F, Angelini J, Taurian T, Tonelli ML & Fabra A (2009) Endophytic occupation of peanut root nodules by opportunistic *Gammaproteobacteria*. *Syst Appl Microbiol* **32**: 49 – 55.

Iglesias O, Rivas R, García-Fraile P, Abril A, Mateos PF, Martinez-Molina E & Velázquez E (2007) Genetic characterization of fast growing rhizobia able to nodulate *Prosopis alba* in North Spain. *FEMS Microbiol Lett* **277**: 210 – 216.

Jarvis BDW, van Berkum P, Chen WX, Nour SM, Fernandez MP, Cleyet-Marel JC & Gillis M (1997) Transfer of *Rhizobium loti, Rhizobium huakuii*, *Rhizobium ciceri*, *Rhizobium mediterraneum*, and *Rhizobium tianshanense* to *Mesorhizobium* gen. nov. *Int J Syst Bacteriol* **47**: 895 – 898.

Jordan DC & Allen ON (1974) Family III. *Rhizobiaceae* (Conn, 1938), *Bergey's manual of determinative bacteriology* (8th edn) (Buchanan RE & Gibbons NE, eds), pp. 261-264. The Williams & Wilkins, Baltimore.

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

Jordan DC (1984) Family III. Rhizobiaceae. *Bergey's manual of systematic bacteriology*, Vol. 1 (Krieg NR & Holt JG, eds), pp. 234 – 242. The Williams and Wilkins, Baltimore.

Jukes TH & Cantor CR (1969) Evolution of protein molecules. (Munro HN, ed), Academic Press, New York.

Kanso S & Patel BKC (2003) *Microvirga subterranea* gen. nov., sp. nov., a moderate thermophile from a deep subsurface Australian thermal aquifer. *Int J Syst Evol Micr* **53**: 401 – 406.

Keatinge JDH, Materon LA, Beck DP, Yurtsever N, Karuc K & Altuntas S (1995) The role of rhizobial biodiversity in legume crop productivity in the West Asian highlands. *Exp Agr* **31**: 485 – 491.

Khurana AS & Sharma P (1995) Variety and Rhizobium strain interactions in lentil. *Lens Newsletter* **22**: 34 – 36.

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Mol Biol Evol* **16**: 111 – 120.

Kitching IJ, Forey PL, Humphries CJ & Williams DM (1998) Cladistics (2nd edn): *The theory and practice of parsimony analysis*. Oxford University Press.

Knösel DH (1984) Genus: *Phyllobacterium*. (Krieg NR & Holt, JG, eds) *Bergey's manual of systematic bacteriology*, Vol. 1, pp. 254 – 256. The Williams & Wilkins, Baltimore.

Konstantinidis KT, Ramette A & Tiedje JM (2006) Toward a more robust assessment of intraspecies diversity, using fewer genetic markers. *Appl Environ Microb* **72**: 7286 – 7293.

Kouchi H, Imaizumi-Anraku H, Hayashi M, Hakoyama T, Nakagawa T, Umehara Y, Suganuma N & Kawaguchi M (2010) How many peas in a pod? legume genes responsible for mutualistic symbioses underground. *Plant Cell Physiol* **51**: 1381 – 1397.

Kumar S & Filipski A (2007) Multiple sequence alignment: in pursuit of homologous DNA positions. *Genome Res* **17**: 127 – 132.

Ladizinsky G (1979) The origin of lentil and its wild genepool. *Euphytica* **28**: 179 – 187.

Laguerre G, Bardin M & Amarger N (1993) Isolation from soil of symbiotic and nonsymbiotic *Rhizobium leguminosarum* by DNA hybridization. *Can J Microbiol* **39**: 1142 – 1149.

Laguerre G, Depret G, Bourion V & Duc G (2007) *Rhizobium leguminosarum* bv. *viciae* genotypes interact with pea plants in developmental responses of nodules, roots and shoots. *New Phytol* **176**: 680 – 690.

Laguerre G, Geniaux E, Marzurier S, Casartelii R & Amarger N (1992) Conformity and diversity among field isolates of *R. leguminosarum* bv. *viceae*, bv. *trifolii*, and bv. *phaseoli* revealed by DNA hybridization using chromosome and plasmid probes *Can J Microbiol* **39**: 412 – 419.

Laguerre G, Louvrier P, Allard MR & Amarger N (2003) Compatibility of rhizobial genotypes within natural populations of *Rhizobium leguminosarum* biovar *viciae* for nodulation of host legumes. *Appl Environ Microb* **69**: 2276 – 2283.

Laguerre G, Mavingui P, Allard MR, Charnay MP, Louvrier P, Mazurier SI, Ligottier-Gois L & Amarger N (1996) Typing of rhizobia by PCR DNA fingerprinting and PCR-restriction fragment length polymorphism analysis of chromosomal and symbiotic gene regions: application to *Rhizobium leguminosarum* and its different biovars. *Appl Environ Microb* **62**: 2029 – 2036.

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

Laguerre G, Nour SM, 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.

Lan R & Reeves P (2001) When does a clone deserve a name? A perspective on bacterial species based on population genetics. *Trends Microbiol* **9**: 419 – 424.

Land AH & Doig AG (1960). An automatic method of solving discrete programming problems". *Econometrica* **28:** 497 – 520

Lavin M, Herendeen PS & Wojciechowski MF (2005) Evolutionary rates analysis of *Leguminosae* implicates a rapid diversification of lineages during the tertiary. *Syst Biol* **54**: 575 – 594.

Lev-Yadun S, Gopher A & Abbo S (2000) The cradle of agriculture. *Science* **288**: 1602 – 1603.

Li JH, Wang ET, Chen WF & Chen WX (2008) Genetic diversity and potential for promotion of plant growth detected in nodule endophytic bacteria of soybean grown in Heilongjiang province of China. *Soil Biol Biochem* **40**: 238 – 246.

Lie TA, Goktan M, Engin M, Pijnenborg J & Anlarsal E (1987) Co-evolution of the legume – rhizobium association. *Plant Soil* **100**: 171 – 181.

Ludwig W (2010) Molecular phylogeny of microorganisms: Is rRNA still a useful marker? *Molecular phylogeny of microorganisms* (Oren A & Papke RT, eds), pp. 65 – 83. Caister Academic Press, Norfolk.

Ludwig W & Klenk HP (2005) Overview: a phylogenetic backbone and taxonomic framework for prokaryotic systematic. *Bergey's manual of systematic bacteriology* Vol. 2 (Brenner DJ, Krieg NR & Staley JT, eds), pp. 49 – 65. Springer, New York.

Maddison WP, Donoghue MJ & Maddison DR (1984) Outgroup analyses and parsimony. *Syst Zool* **33**: 83 – 103.

Maiden MC (2006) Multilocus sequence typing of bacteria. *Ann Rev Microbiol* **60**: 561 – 588.

Makkar NS & Casida LE (1987) *Cupriavidus necator* gen. nov., sp. nov.: a non obligate bacterial predator of bacteria in soil. *Int J Syst Bacteriol* **37**: 323 – 326.

Martens M, Dawyndt P, Coopman R, Gillis M, Vos PD & Willems A (2008) Advantages of multilocus sequence analysis for taxonomic studies: a case study using

housekeeping genes in the genus *Ensifer* (including former *Sinorhizobium)*. *Int J Syst Evol Micr* **58**: 200 – 214.

Masson-Boivin C, Giraud E, Perret X & Batut J (2009) Establishing nitrogen-fixing symbiosis with legumes: how many rhizobium recipes? *Trends Microbiol* **17**: 458 – 466.

Mathews S & Donoghue MJ (1999) The root of angiosperm phylogeny inferred from duplicate phytochrome genes. *Science* **286**: 947 – 950.

Mau B, Newton M & Larget B (1999) Bayesian phylogenetic inference via Markov chain Mont Carlo methods. *Biometrics* 55: 1 – 12

Mayr E (1963) *Animal species and evolution.* Harvard University Press; London: Oxford University Press.

McKey D (1994) Legumes and nitrogen: The evolutionary ecology of a nitrogen-demanding lifestyle. *Advances in legume systematics* (Sprent JI & McKey D, eds), Royal Botanic Gardens, Kew, UK.

McNeil DL & Materne M (2007) Rhizobium management and nitrogen fixation *Lentil: an ancient crop for Modern times* (Yadav SS, McNeil DL & Stevenson PC, eds), pp. 127 – 143. Springer, Dordrecht, The Netherlands.

Mergaert P, Van Montagu M & Holsters M (1997) Molecular mechanisms of Nod factor diversity. *Mol Microbiol* **25**: 811 – 817.

Milkman R & Bridges MM (1990) Molecular evolution of the *E. coli* chromosome III Clonal frames. *Genetics* **126**: 505 – 517.

Moawad H, Badr El-Din SMS & Abdel-Aziz RA (1998) Improvement of biological nitrogen fixation in Egyptian winter legumes through better management of *Rhizobium*. *Plant Soil* **204**: 95 – 106.

Moawad H & Beck DP (1991) Some characteristics of *Rhizobium leguminosarum* isolates from un-inoculated field-grown lentil. *Soil Biol Biochem* **23**: 933 – 937.

Moron B, Soria-Diaz ME, Ault J, Verrios G, Noreen S, Rodriguez-Navarro D, Gil-Serrano A, Thomas-Oates J, Megias M & Sousa C (2005) Low pH changes the profile of nodulation factors produced by *Rhizobium tropici* CIAT899. *Chem Biol* **12**: 1029 – 1040.

Moschetti G, Peluso A, Protopapa A, Anastasio M, Pepe O & Defez R (2005) Use of nodulation pattern, stress tolerance, *nodC* gene amplification, RAPD-PCR and RFLP-16S rDNA to discriminate genotypes of *Rhizobium leguminosarum* biovar *viciae*. *Syst Appl Microbiol* **28**: 619 – 631.

Moulin L, Chen W-M, Béna G, Dreyfus B & Boivin-Masson C (2002) Rhizobia: The family is expanding. *Nitrogen fixation: global perspectives* (Finan T, O'Brian M, Layzell D, Vessey K & Newton W, eds), pp. 61 – 65. CAB International, Wallingford .

Mutch LA & Young JPW (2004) Diversity and specificity of *Rhizobium leguminosarum* biovar *viciae* on wild and cultivated legumes. *Mol Ecol* **13**: 2435 – 2444.

Mwangi SN, Karanja NK, Boga H, Kahindi JHP, Muigai A, Odee D & Mwenda GM (2011) Genetic diversity and symbiotic efficiency of legume nodulating bacteria from different land use system in Taita Taveta, Kenya. *Trop subtrop Agro-ecosystems* **13**: 109 – 118.

Nakagawa Y, Sakane T & Yokota A (1996) Transfer of "*Pseudomonas riboflavina*" (Foster 1944), a gram-negative, motile rod with long-chain 3-hydroxy fatty acids, to *Devosia riboflavina* gen. nov., sp. nov., nom. rev. *Int J Syst Bacteriol* **46**: 16 – 22.

Naser SM, Thompson FL, Hoste B, Gevers, D, Dawyndt P, Vancanneyt M & Swings J (2005) Application of multilocus sequence analysis (MLSA) for rapid identification of *Enterococcus* species based on *rpoA* and *pheS* genes. *Microbiology* **151**: 2141 – 2150.

Narra H & Ochman H (2006) Of what use is sex to bacteria? *Curr Biol* **16**: 705 – 710.

Nautiyal CS, Srivastava S & Chauhan PS (2008) Rhizosphere colonization: molecular determinant from plant-microbe coexistence perspective. *Molecular mechanisms of plant and microbe coexistence* (Nautiyal CS & Dion P, eds), pp. 99 – 123 . Springer-Verlag, Heidelberg.

Nei M (1987) *Molecular evolutionary genetics*. Columbia University Press, New York.

Nei M & Kumar S (2000) *Molecular evolution and phylogenetics* Oxford University Press, New York.

Nene YL (2006) Indian pulse through the millennia. *Asian Agri-History* **10**: 179 – 202

Ngom A, Nakagawa, Y, Sawada, H, *et al*. (2004). A novel symbiotic nitrogen-fixing member of the Ochrobactrum clade isolated from root nodules of *Acacia mangium*.
J Gen Appl Microbiol **50**:17 – 27.

Nobbe F & Hiltner L (1896) Inoculation of the soil for cultivating leguminous plants. *US Patent* 570813.

Ochman H, Lawrence JG & Groisman EA (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**: 299 – 304.

Ollivier J, We ST, Bannert A, Hai B, Kastl EM, Meyer A, Su MX, Kleineidam K & Schloter M (2011) Nitrogen turnover in soil and global change. *FEMS Microbiol Ecol* **78**: 3 – 16.

Oren A (2010) Concepts about phylogeny of microorganisms - an historical overview *Molecular Phylogeny of Microorganisms* (Oren A & Papke RT, eds), pp. 1 – 21. Caister Academic Press, Norfolk.

Palys T, Nakamura LK & Cohan FM (1997) Discovery and classification of ecological diversity in the bacterial world: the role of DNA sequence data. *Int J Syst Bacteriol* **47**: 1145 – 1156.

Patt TE, Cole GC & Hanson RS (1976) *Methylobacterium*, a new genus of facultatively methylotrophic bacteria. *Int J Syst Bacteriol* **26**: 226 – 229.

Perez-Ramírez ON, Rogel MA, Wang E, Castellanos ZJ & Martínez-Romero E (1998) Seeds of *Phaseolus vulgaris* bean carry *Rhizobium etli*. *FEMS Microbiol Ecol* **26**: 289 – 296.

Perret X, Staehelin C & Broughton WJ (2000) Molecular basis of symbiotic promiscuity. *Microbiol Mol Biol R* **64**: 180 – 201.

Pritchard J, Stephens M & Donnelly P (2000) Inference of population structure using multilocus genotype data *Genetics* **155**: 945 – 959.

Rahman MM, Bakr MA, Mia MF, Idris KM, Gowda CLL, Kumar J, Deb UK, Malek MA & Sobhan A (2009) Legumes in Bangladesh. ICARDA, Patancheru, Andhra Pradesh, India.

Ramírez-Bahena M, García-Fraile P, Peix A, Valverde A, Rivas R, Igual, J.M. , Mateos PF, Martínez-Molina E & Velázquez E (2008) Revision of the taxonomic status of the species *Rhizobium leguminosarum* (Frank 1879) Frank 1889AL, *Rhizobium phaseoli* Dangeard 1926AL and *Rhizobium trifolii* Dangeard 1926AL. *R. trifolii* is a later synonym of *R. leguminosarum*. Reclassification of the strain *R. leguminosarum* DSM 30132 (=NCIMB 11478) as *Rhizobium pisi* sp. nov. *Int J Syst Evol Microbiol* **58**: 2484-2490.

Rannala B & Yang ZH (1996) Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. *J Mol Evol* **43:** 304 – 311

Rashid MH, Satter MA, Uddin MI & Young JPW (2009) Molecular characterization of symbiotic root nodulating rhizobia isolated from lentil (*Lens culinaris*). *EJEAFChe* **8**: 602 – 612.

Rashid MH, Schäfer H, Gonzalez J & Wink M (2012) Genetic diversity of rhizobia nodulating lentil (*Lens culinaris*) in Bangladesh. *Syst Appl Microbiol* **35**: 98 – 109.

Redfield R (2001) Do bacteria have sex? *Nat Rev Genet* **2**: 634 – 639.

Risal CP, Djedidi S, Dhaka D, Ohkama-Ohtsu N, Sekimoto H & Yokoyama T (2012) Phylogenetic diversity and s ymbiotic functioning in mungbean (*Vigna radiata*) bradyrhizobia from contrast argo-ecological region of Nepal. *Syst Appl Microbiol* **35**: 45 – 53.

Rivas R, García-Fraile P & Velázquez E (2009) Taxonomy of bacteria nodulating legumes. *Microbiol Insights* **2**: 51 – 69.

Rodríguez F, Oliver JF, Marín A & Medina JR (1990) The general stochastic model of nucleotide substitution. *J Theor Biol* **142**: 485 – 501.

Rogel MA, Ormeno-Orrillo E & Martinez Romero E (2011) Symbiovars in rhizobia reflect bacterial adaptation to legumes. *Syst Appl Microbiol* **34**: 96 – 104.

Ronquist F & Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**: 1572 – 1574.

Rosselló-Mora R & Amann R (2001) The species concept for prokaryotes. *FEMS Microbiol Rev* **25**: 39 – 67.

Rozas J, Sanchez-DelBarrio JC, Messeguer X & Rozas R (2010) DnaSP: DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**: 2496 – 2497.

Saha BK, Chowdhury MAH & Rashid MH (2008) Effect of rhizobia and host cultivar on some biochemical constituents and yield of lentil. *Int J Biores* **4**: 100 – 107.

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

Santillana N, Ramírez-Bahena MH, García-Fraile P, Velázquez E & Zuniga D (2008) Phylogenetic diversity based on rrs, *atpD, recA* genes and 16S – 23S intergenic sequence analyses of rhizobial strains isolated from *Vicia faba* and *Pisum sativum* in Peru. *Arch Microbiol* **189**: 239 – 247.

Sarker A & Erskine W (2006) Recent progress in the ancient lentil. *J Agr Sci* **144**: 19 – 29.

Satter MA (2001) Soil Microbiology. *Agricultural research in Bangladesh in the 20th centtury* (Mian MAW, Maniruzzaman FM, Satter MA, Miah MAA, Paul SK & Haque KR, eds), pp. 117 – 123. Bangladeh agricultural research council & Bangladesh academy of agriculture, Dhaka.

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

Schlüter PM & Harris SA (2006) Analysis of multilocus fingerprinting data sets containing missing data. *Mol Ecol notes* **6**: 569 – 572.

Schmidt HA, Petzold E, Vingron M & von Haeseler A (2003) Molecular phylogenetics: parallelized parameter estimation and quartet puzzling. *J Parallel Distr Com* **63**: 719 – 727.

Scholla MH & Elkan GH (1984) *Rhizobium fredii* sp.nov., a fast-growing species that effectively nodulates soybeans. *Int J Syst Bacteriol* **34**: 484 – 486.

Schuster ML & Coyne DP (1974) Survival mechanisms of phytopathogenic bacteria *Ann Rev phytopathol* **12**: 199 – 221.

Segovia L, Young JPW & Martínez-Romero E (1993) Reclassification of American *Rhizobium leguminosarum* biovar *phaseoli* type I strains as *Rhizobium etli* sp. nov. *Int J Syst Bacteriol* **43**: 374 – 377.

Shandu JS & Singh S (2007) History and origin. Lentil: an ancient crop for modern times. (Shyam SY, David MN & Stevenson PC, eds), pp. 1 – 9. Springer, Dordrecht.

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

Silva C, Vinuesa P, Eguiarte L, Souza V & Martínez-Romero E (2005) Evolutionary genetics and biogeographic structure of *Rhizobium gallicum sensu lato*, a widely distributed bacterial symbiont of diverse legumes. *Mol Ecol* **14**: 4033 – 4050.

Simmons MP, Bailey DC & Nixon KC (2000) Phylogeny reconstruction using duplicate genes. *Mol Biol Evol* **17**: 469 – 473.

Skerman VBD, Mcgowan V & Sneath PHA (1980) Approved lists of bacterial names. *Int J Syst Bacteriol* **30**: 225 – 420.

Slattery JF & Pearce D (2002) Development of elite inoculant Rhizobium strains in southeastern Australia. (Herridge D, ed), pp. 86 – 94. ACIAR, Australia.

Smith AB (1994) Rooting molecular trees: problem and strategies. *Biol J Linn Soc* **51**: 279 – 292.

Somasegaran P & Hoben HJ (1994) *Handbook for rhizobia: methods in legume-rhizobium technology.* Springer, Heidelberg.

Sonnante G, Hammer K & Pignone D (2009) From the cradle of agriculture a handful of lentils:history of domestication. *Rendiconti Lincei* **20**: 21 – 37.

Sprent JI (1994) Evolution and diversity in the legume-rhizobium symbiosis: chaos theory? *Plant Soil* **161**: 1 – 10.

Sprent JI (2001) *Nodulation in Legumes.* Royal Botanic Gardens, UK.

Stepkowski T, Moulin L, Krzyzanska A, McInnes A, Law IJ & Howieson J (2005) European origin of *Bradyrhizobium* populations infecting lupins and serradella in soils of Western Australia and South Africa. *Appl Environ Microb* **71**: 7041 – 7052.

Strain SR, Whittam TS & Bottomley PJ (1995) Analysis of genetic structure in soil populations of *Rhizobium leguminosarum* recovered from the USA and the UK. *Mol Ecol* **4**: 105 – 114.

Swofford DL, Olsen GJ, Waddell P & HilliS DM (1996) Phylogenetic inference. (Hillis DM, Moritz C & Mable BK, eds), pp. 407 – 514. Sinauer Associates, Sunderland, Massachusetts.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M & Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**: 2731 – 2739.

Tavaré S (1986) Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on Mathematics in the Life Sciences* (American Mathematical Society) **17**: 57 – 86.

Tegegn ND (2006) Genetic diversity and characterization of indigenous *Rhizobium leguminosarum* biovar viciae isolates of cool-season food legumes growth in the highlands of Ethiopia. Thesis University of Putra, Malaysia.

Tian CF, Young JPW, Wang ET, Tamimi SM & Chen WX (2010) Population mixing of *Rhizobium leguminosarum* bv. *viciae* nodulating *Vicia faba*: the role of recombination and lateral gene transfer. *FEMS Microbiol Ecol* **73**: 563 – 576.

Tian CF, Zhoub YJ, Zhanga YM *et al.,* (2012) Comparative genomics of rhizobia nodulating soybean suggests extensive recruitment of lineage-specific genes in adaptations. *P Natl Acad Sci USA*.

Toklu F, Karakoy T, Hakli E, Bicer T, Brandolini A, Kilian B & Özkan H (2009) Genetic variation among lentil (*Lens culinaris* Medik) land races from Southeast Turkey. *Plant Breeding* **128**: 178 – 186.

Unkovich MJ & Pate JS (2000) An appraisal of recent field measurements of symbiotic N_2 fixation by annual legumes. *Field Crops Res* **65**: 211 – 228.

Urakami T, Araki H, Oyanagi H, Suzuki KI & Komagata K (1992) Transfer of *Pseudomonas aminovorans* (den Dooren de Jong 1926) to *Aminobacter* gen. nov. as *Aminobacter aminovorans* comb. nov. and description of *Aminobacter aganoensis* sp. nov. and *Aminobacter niigataensis* sp. nov. *Int J Syst Bacteriol* **42**: 84 – 92.

Valverde A, Igual JM, Peix A, Cervantes E & Velazquez E (2006) *Rhizobium lusitanum* sp. nov. a bacterium that nodulates *Phaseolus vulgaris*. *Int J Syst Evol Micr* **56**: 2631 – 2637.

Van de Peer Y (2003) Phylogeny inference based on distance methods. (Salemi M & Vandamme AM, eds), pp. 101 – 119. Cambridge University Press, Cambridge.

Vandamme AM (2009) Basic concept of molecular evolution. *The phylogenetic handbook: A practical approach to the analysis and hypothesis testing* (Lamey P, Salemi M & Vandamme AM, eds), pp. 3 – 29. Cambridge University press, Cambridge.

Versalobic J, Koeuth T & Lupski JR (1991) Distribution of repetitive DNA sequences in eubacteria and application to fingerprinting of bacterial genome. *Nucleic Acids Research* **19**: 6823 – 6831.

Vessey JK (2004) Benefits of inoculating legume crops with rhizobia in the Northern Great Plains. *Crop Management,* doi: 10.1094 / CM-2004-0301-04-RV.

Vincent JM (1970) A Manual for the Practical Study of Root-Nodule Bacteria, Blackwell, Oxford.

Vinuesa P, Silvaa C, Wernerb D & Martı´nez-Romero E (2005) Population genetics and phylogenetic inference in bacterial molecular systematics: the roles of migration and recombination in *Bradyrhizobium* species cohesion and delineation. *Mol Phylo Evol* **34**: 29 – 54.

Wang ET, Rogel MA & García-de los Santos A (1999) *Rhizobium etli* bv. *mimosae*, a novel biovar isolated from *Mimosa affinis*. *Int J Syst Evol Micr* **49**: 1479 – 1491.

Wang ET, van Berkum P, Beyene D, Sui XH, Dorado O, Chen WX & Martinez-Romero E (1998) *Rhizobium huautlense* sp. nov., a symbiont of *Sesbania herbacea* that has a close phylogenetic relationship with *Rhizobium galegae*. *Int J Syst Evol Micr* **48**: 687 – 699.

Wang L, Gruber S & Claupein W (2012) Effect of sowing date and variety on yield and weed populations in a lentil-barley mixture. *J Agril Sci* **In press**: 1 – 10.

Weisburg WG, Barns SM, Pelletier DA & Lane D J (1991) 16S ribosomal DNA amplification for phylogenetic study. *J Bacteriol* **173**: 697 – 703.

Wernegreen JJ, Harding EE & Riley MA (1997) Rhizobium gone native: unexpected plasmid stability of indigenous *R. leguminosarum*. *P Natl Acad Sci USA* **94**: 5483 – 5488.

Wernegreen JJ & Riley MA (1999) Comparison of the evolutionary dynamics of symbiotic and housekeeping loci: a case for the genetic coherence of rhizobial lineages. *Mol Biol Evol.* **16**: 98 – 113.

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

Wielbo J, Marek-Kozaczuk M, Kidaj D & Skorupska A (2011) Competitiveness of *Rhizobium leguminosarum* bv. *trifolii* Strains in mixed inoculation of clover (*Trifolium pratense*). *Pol J Microbiol* **60**: 43 – 49.

Willems A (2006) The taxonomy of rhizobia: an overview. *Plant Soil* **287**: 3 – 14.

Wink M (2007) Systematics. *Raptor Research and Management Technique* (Bird DM & Bildstein KL, eds), pp. 57 – 72. Hancock House Publishers, Surrey.

Woese CR & Fox GE (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci USA* **74**: 5088 – 5090.

Woese CR, Stackebrandt E, Weisburg WG *et al.,* (1984) The phylogeny of purple bacteria: The alpha subdivision. *System Appl Microbiol* **5**: 315 – 326.

Wolde-Meskel E, Terefework Z, Frostegård A & Lindstrom K (2005) Genetic diversity and phylogeny of rhizobia isolated from agro-forestry legume species in southern Ethiopia *Int J Syst Evol Micr* **55**: 1439 – 1452.

Yabuuchi E, Kosako Y, Oyaizu H, Yana I, Hotta H, Hashimoto Y, Ezaki T & Arakawa M (1992) Proposal of *Burkholderia* gen. nov. and transfer of seven species of the genus *Pseudomonas* homology group II to the new genus, with the type species *Burkholderia cepacia* (Palleroni and Holmes 1981) comb. nov. *Microbiol. Immunology* **36**: 1251 – 1275.

Young JM, Kuykendall LD, Martinenez-Romero E, Kerr A & Sawada H (2001) A revission of *Rhizobium* (Frank 1889), with an emended description of the genus , and the inclussion of all species of Agrobacterium (Conn 1942) and *Allorhizobium undicola* (de Lajudie et al.1998) as new combinations: *Rhizobium radiobacter*, *R. rhizogenes, R. rubi, R. undicola* and *R. vitis*. *Int J Syst Evol Micr* **51**: 89 – 103.

Young JPW (1996) Phylogeny and taxonomy of rhizobia. *Plant Soil* **186**: 45 – 52.

Young JPW & Haukka K (1996) Diversity and phylogeny of rhizobia. *New Phytol* **133**: 87 – 94.

Young JPW & Wexler M (1988) Sym plasmid and chromosomal genotypes are correlated in field populations of *Rhizobium leguminosarum*. *J Gen Microbiol* **134**: 2731 – 2739.

Zakhia F, Jeers H, Willems A, Gillis M, Dreyfus B & de Laj udie P (2006) Diverse bacteria associated with root nodules of spontaneous legumes in Tunisia and first report for nifH-like gene within the genera *Microbacterium* and *Starkeya*. *Microbial Ecol* **51**: 375 – 393.

Zeze A, Mutch LA & Young JPW (2001) Direct amplification of *nodD* from community DNA reveals the genetic diversity of *Rhizobium leguminosarum* in soil. *Environ Microbiol* **3**: 363 – 370.

Zhang XX, Kosier B & Priefer UB (2001) Genetic diversity of indigenous *Rhizobium leguminosarum* bv. *viciae* isolates nodulating two different host plant during soil restoration with alfalfa. *Mol Ecol* **10**: 2297 – 2305.

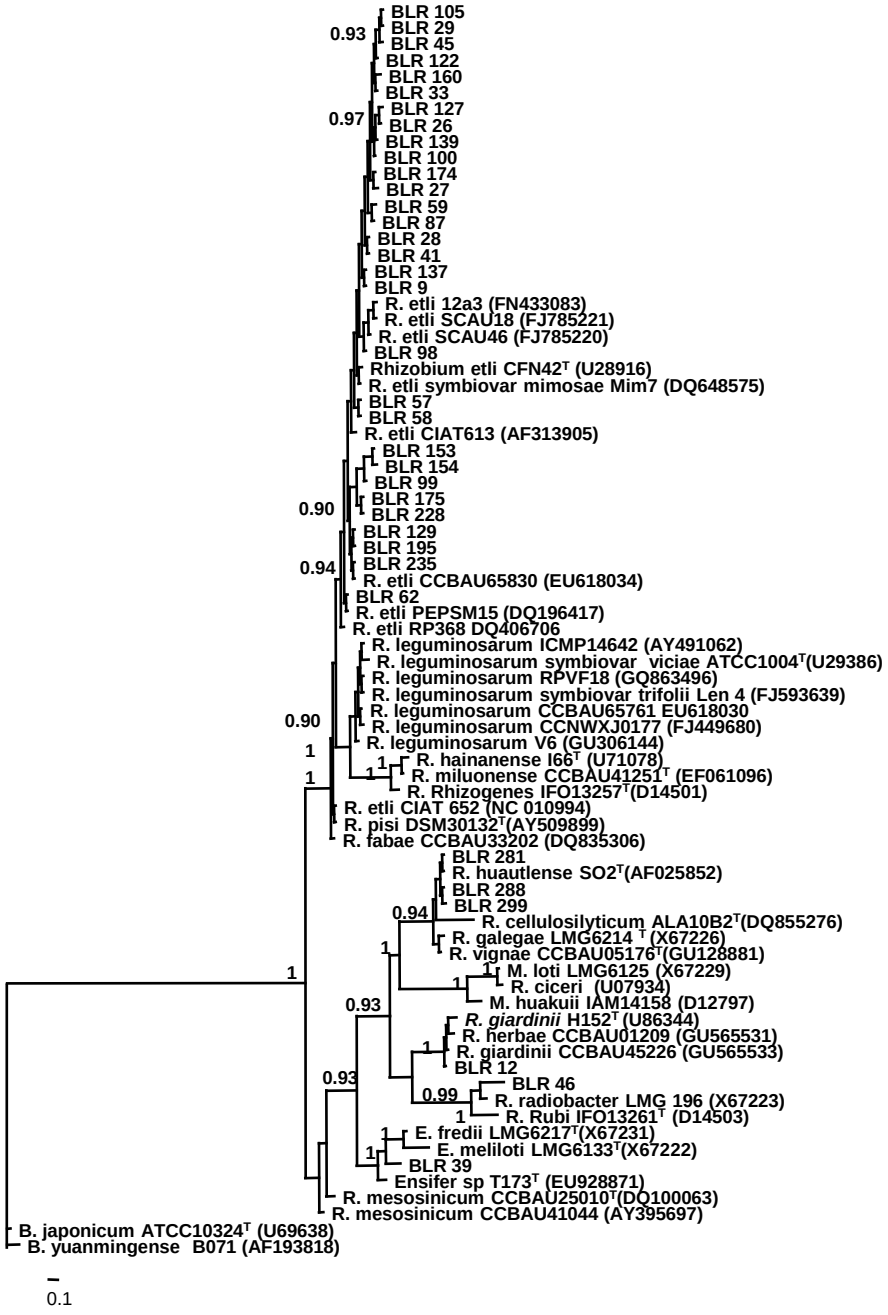
Zohary D & Hopf M (1993) Domestications of plants in the Old World: the origin and spread of cultivated plants in West Africa, Europe and the Nile Valley. Clarendon Press, Oxford.

Zohary D & Hopf M (2000) *The domestication of plants in the Old World.* Oxford University Press.

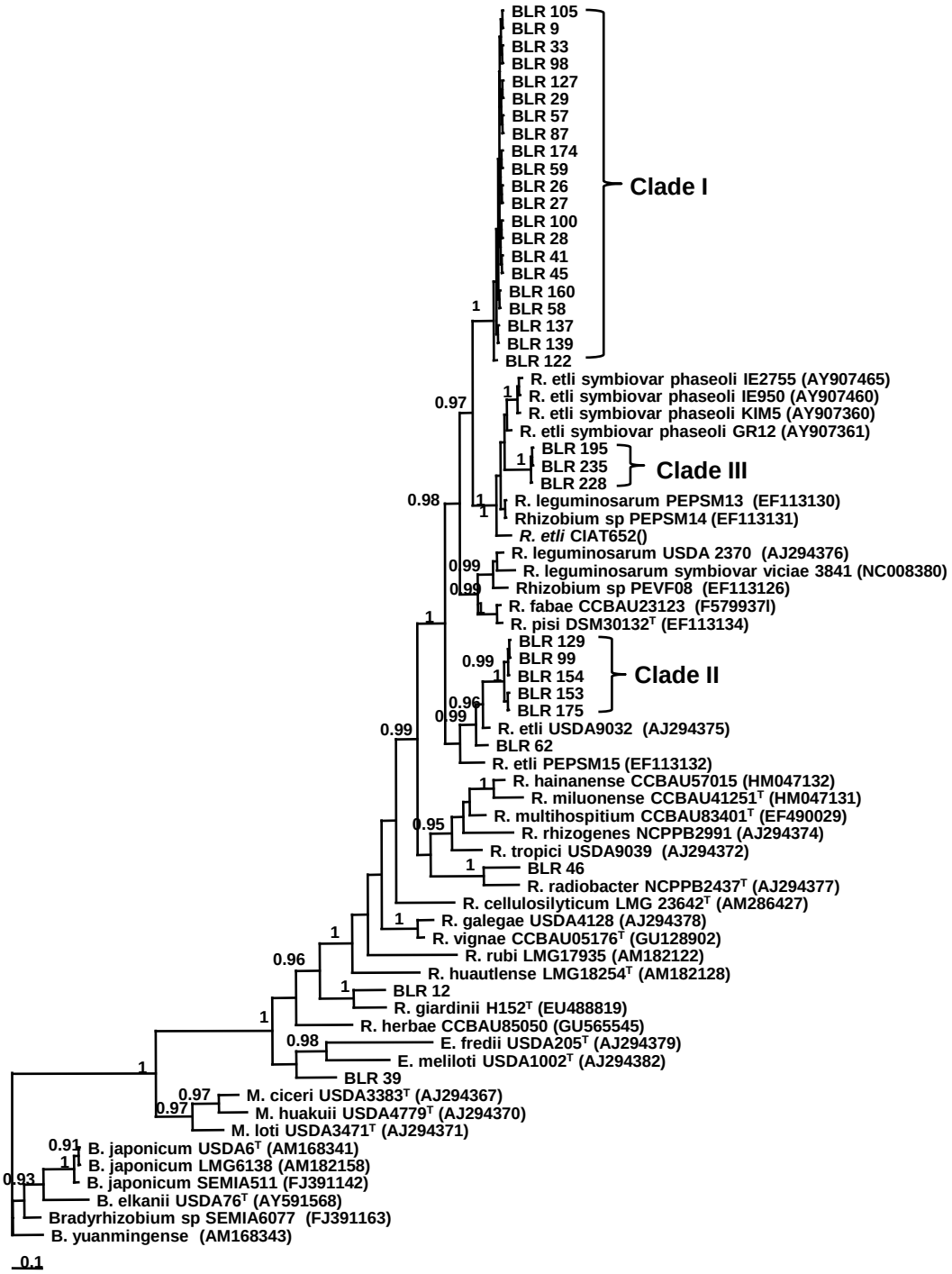
Zuckerkandl E & Pauling L (1965) Molecules as documents of evolutionary history. *J Theor Biol* **8**: 357-366.

Zurkowski W & Lorkiewicz Z (1976) Plasmid deoxyribonucleic acid in *Rhizobium trifolii*. *J Bacteriol* **128**: 481 – 484.

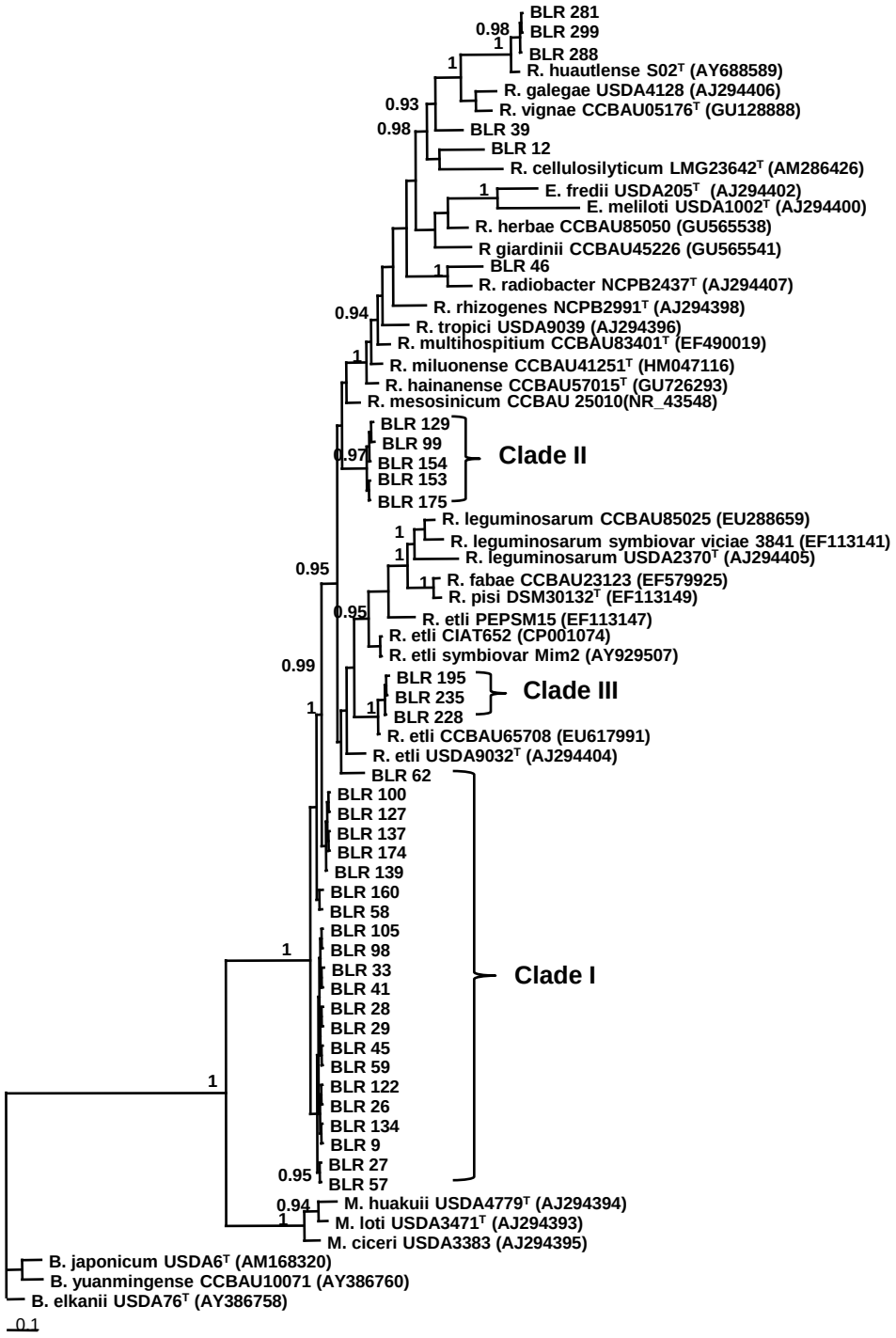
6. Appendixes



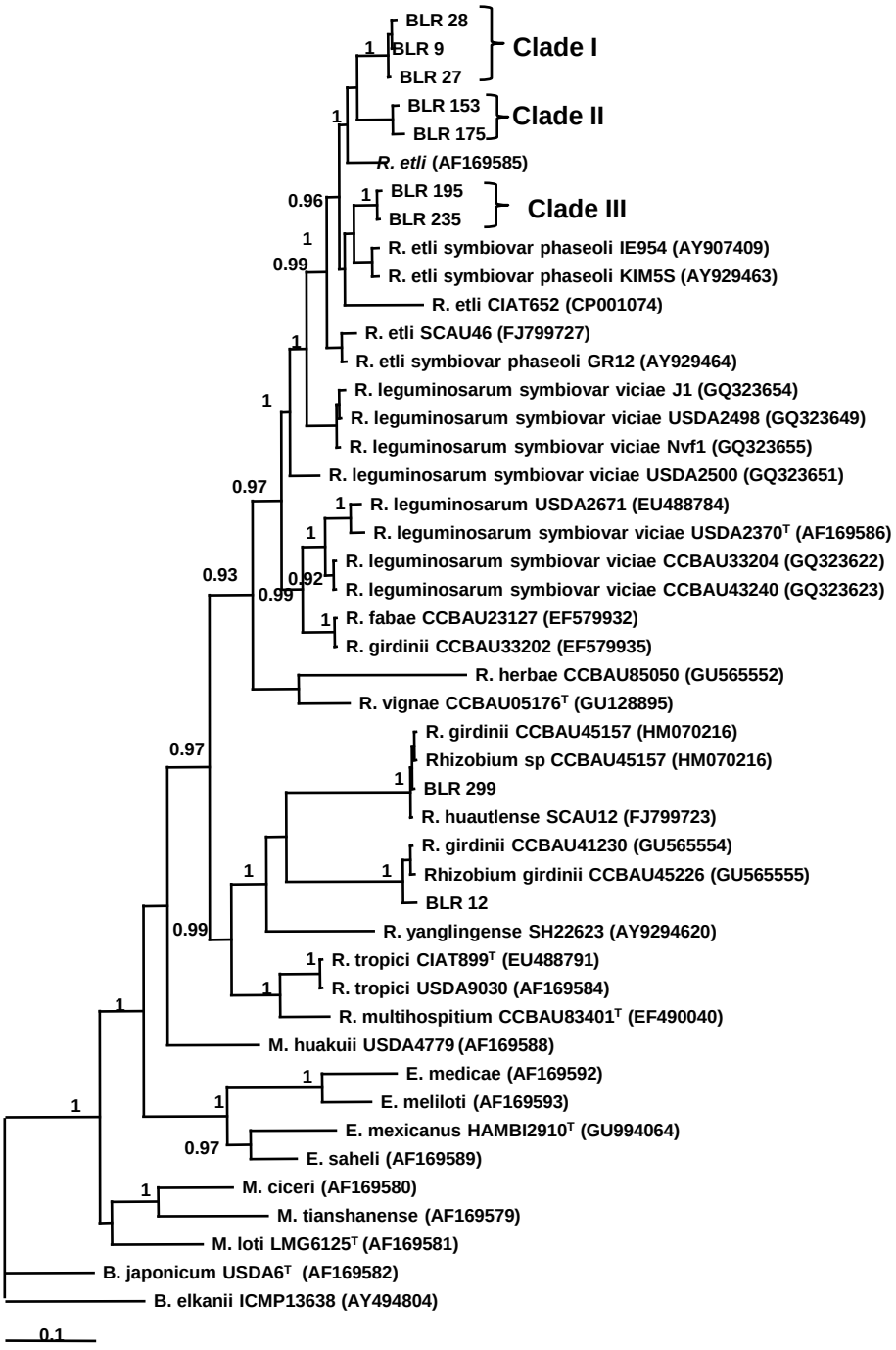
Appendix 1. Bayesian tree based on 16S rRNA gene partial sequences. Posterior probability values shown when ≥ 0.90. Abbreviations used: BLR: Bangladeshi lentil rhizobia., R: *Rhizobium,* E*: Ensifer,* B*: Bradyrhizobium,* M*: Mesorhizobium*



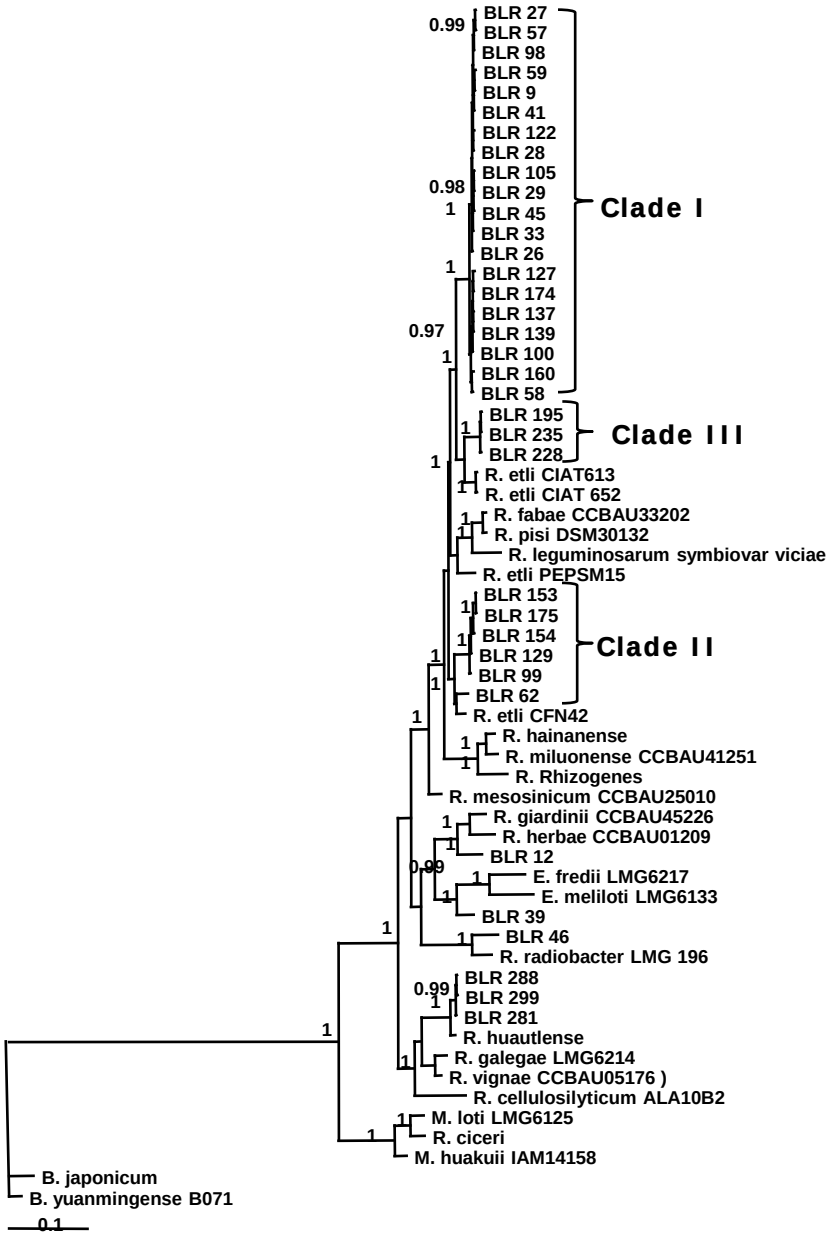
Appendix 2. Bayesian tree based on *recA* gene partial sequences. Posterior probability values shown when ≥ 0.90. Abbreviations used: BLR: Bangladeshi lentil rhizobia., R: *Rhizobium,* E*: Ensifer,* B*: Bradyrhizobium,* M*: Mesorhizobium*



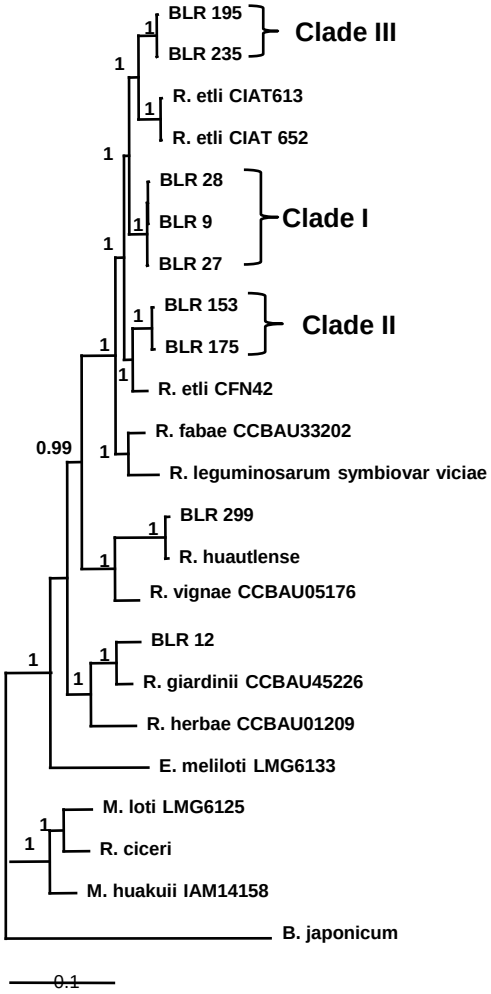
Appendix 3. Bayesian tree based on *atpD* gene partial sequences. Posterior probability values shown when ≥ 0.90. Abbreviations used: BLR: Bangladeshi lentil rhizobia., R: *Rhizobium,* E*: Ensifer,* B*: Bradyrhizobium,* M*: Mesorhizobium*



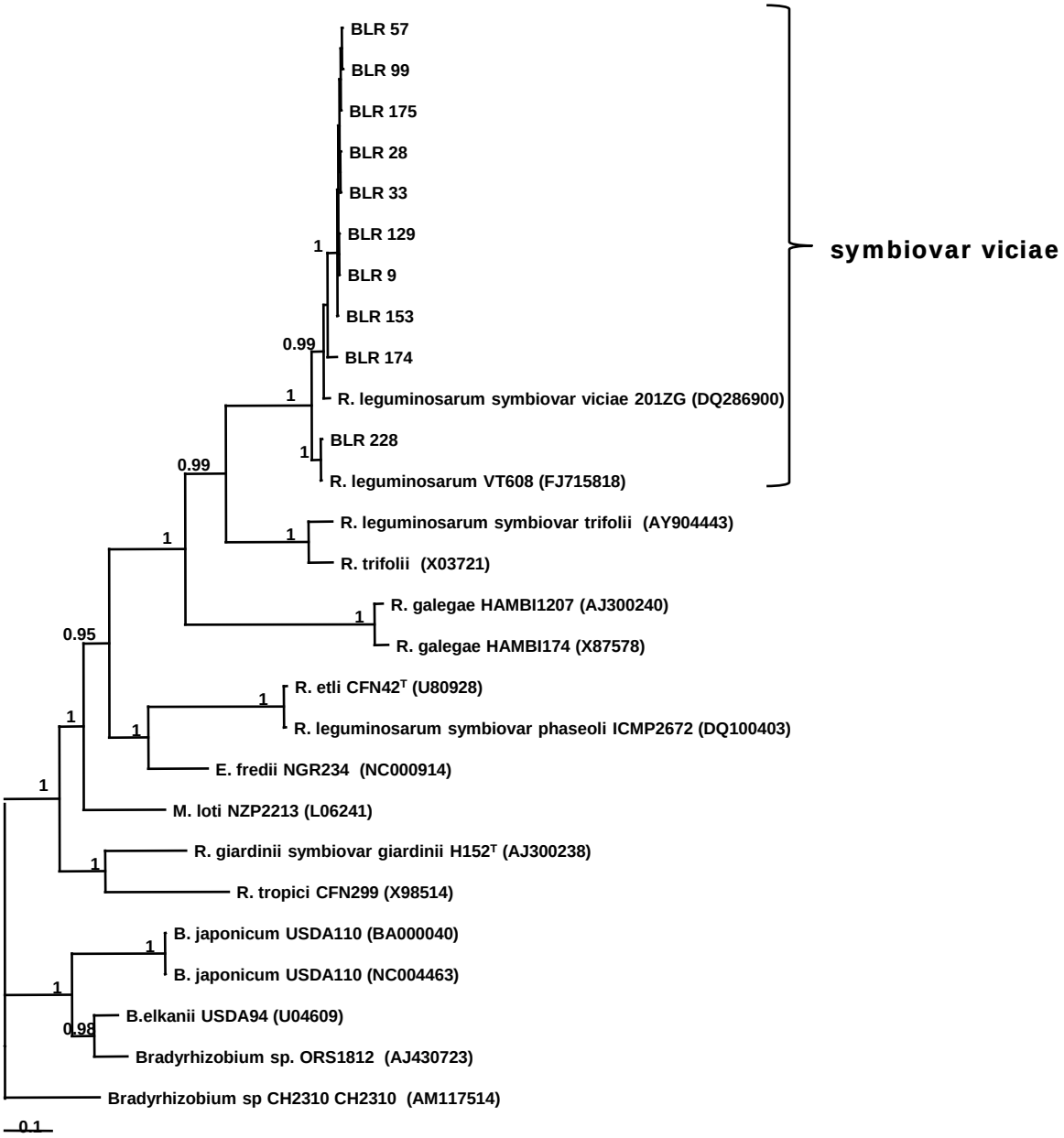
Appendix 4. Bayesian tree based on *glnII* gene partial sequences. Posterior probability values shown when ≥ 0.90. Abbreviations used: BLR: Bangladeshi lentil rhizobia., R: *Rhizobium,* E*: Ensifer,* B*: Bradyrhizobium,* M*: Mesorhizobium*



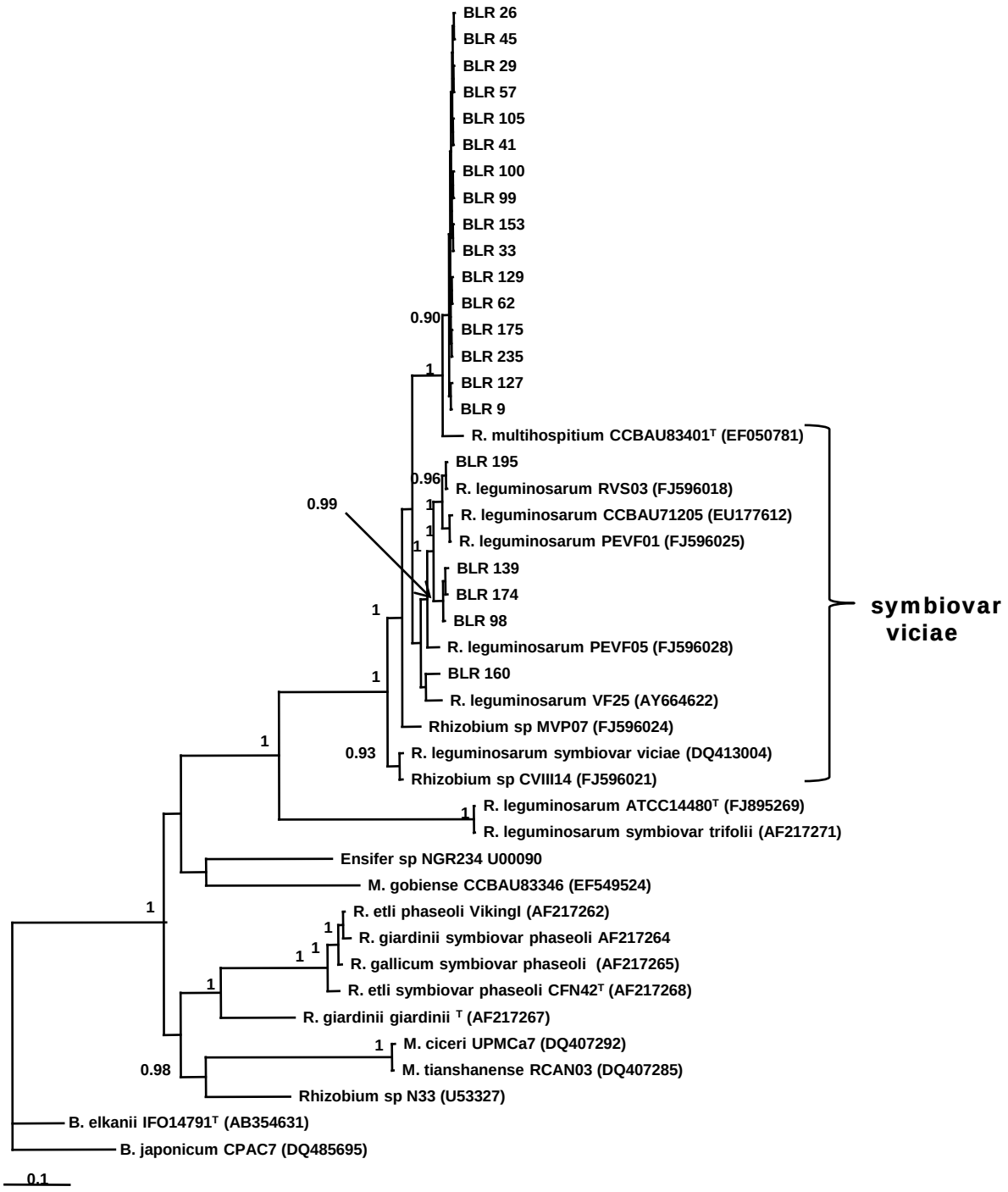
Appendix 5. Bayesian tree based on the concatenated partial sequences of 16S-*atpD-recA* genes. Posterior probability values shown when ≥ 0.90. Abbreviations used: BLR: Bangladeshi lentil rhizobia, R: *Rhizobium,* E*: Ensifer,* B*: Bradyrhizobium,* M*: Mesorhizobium*



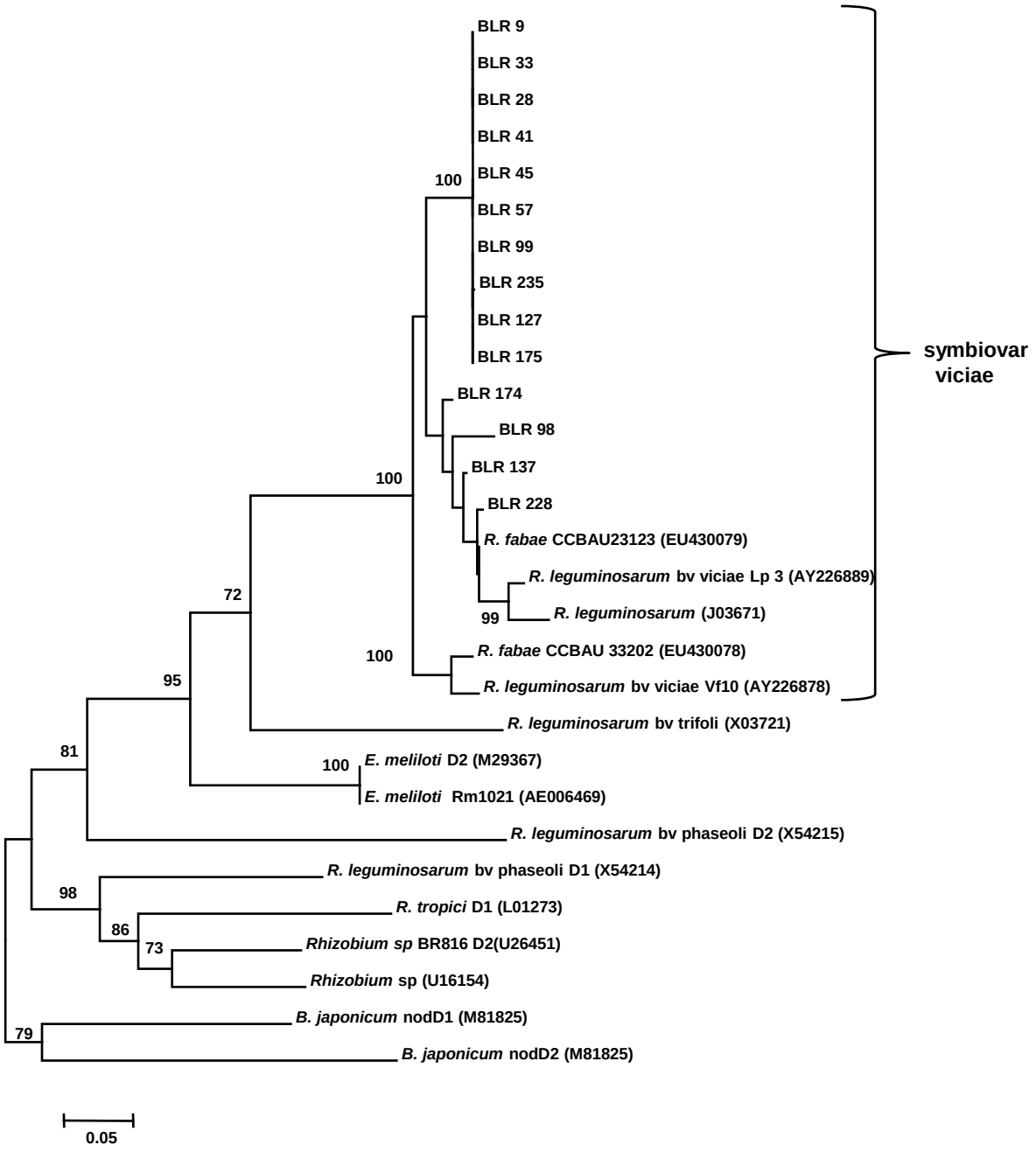
Appendix 6. Bayesian tree based on the concatenated partial sequences of 16S-*atpD-recA-glnII* genes. Posterior probability values shown when ≥ 0.90. Abbreviations used: BLR: Bangladeshi lentil rhizobia., R: *Rhizobium,* E*: Ensifer,* B*: Bradyrhizobium,* M*: Mesorhizobium*



Appendix 7. Bayesian tree based on *nodA* gene partial sequences. Posterior probability values shown when ≥ 0.90. Abbreviations used: BLR: Bangladeshi lentil rhizobia., R: *Rhizobium,* E*: Ensifer,* B*: Bradyrhizobium,* M*: Mesorhizobium*

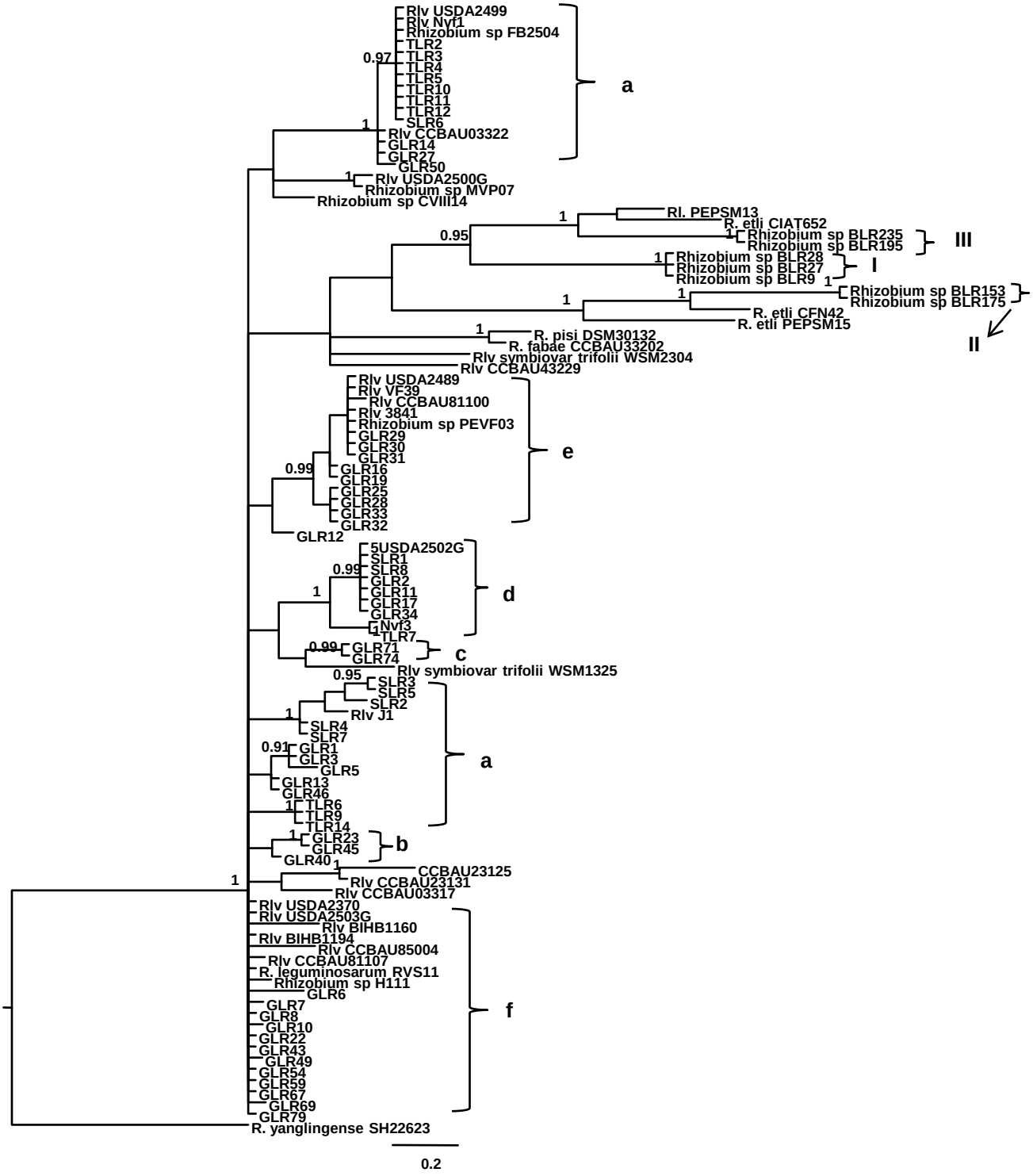


Appendix 8. Bayesian tree based on *nodC* gene partial sequences. Posterior probability values shown when ≥ 0.90. Abbreviations used: BLR: Bangladeshi lentil rhizobia., R: *Rhizobium,* E*: Ensifer,* B*: Bradyrhizobium,* M*: Mesorhizobium*

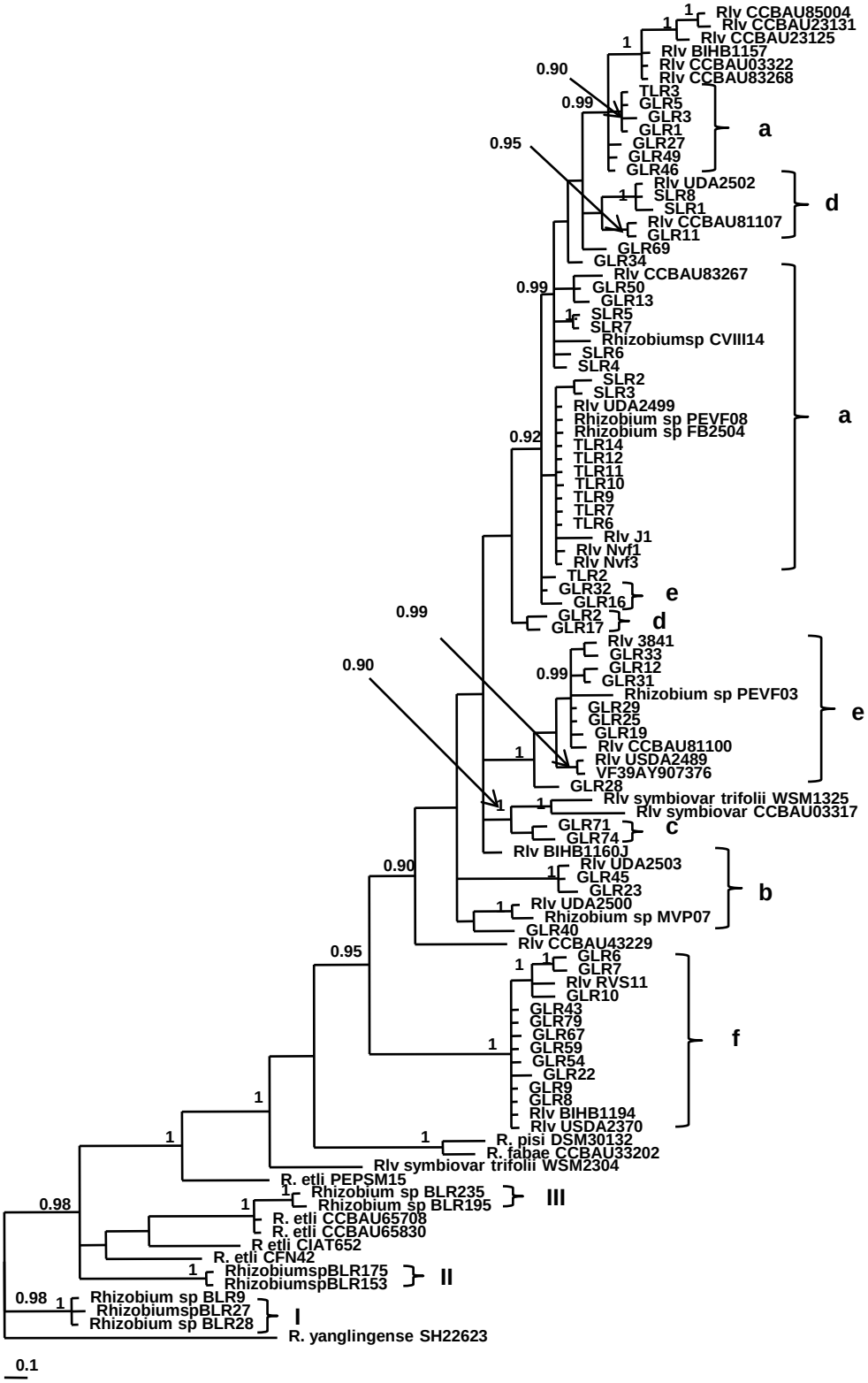


Appendix 9. Bayesian tree based on *nodD* gene partial sequences. Posterior probability values shown when ≥ 0.90. Abbreviations used: BLR: Bangladeshi lentil rhizobia., R*hizobium,* E*: Ensifer,* B*: Bradyrhizobium,* M*: Mesorhizobium*

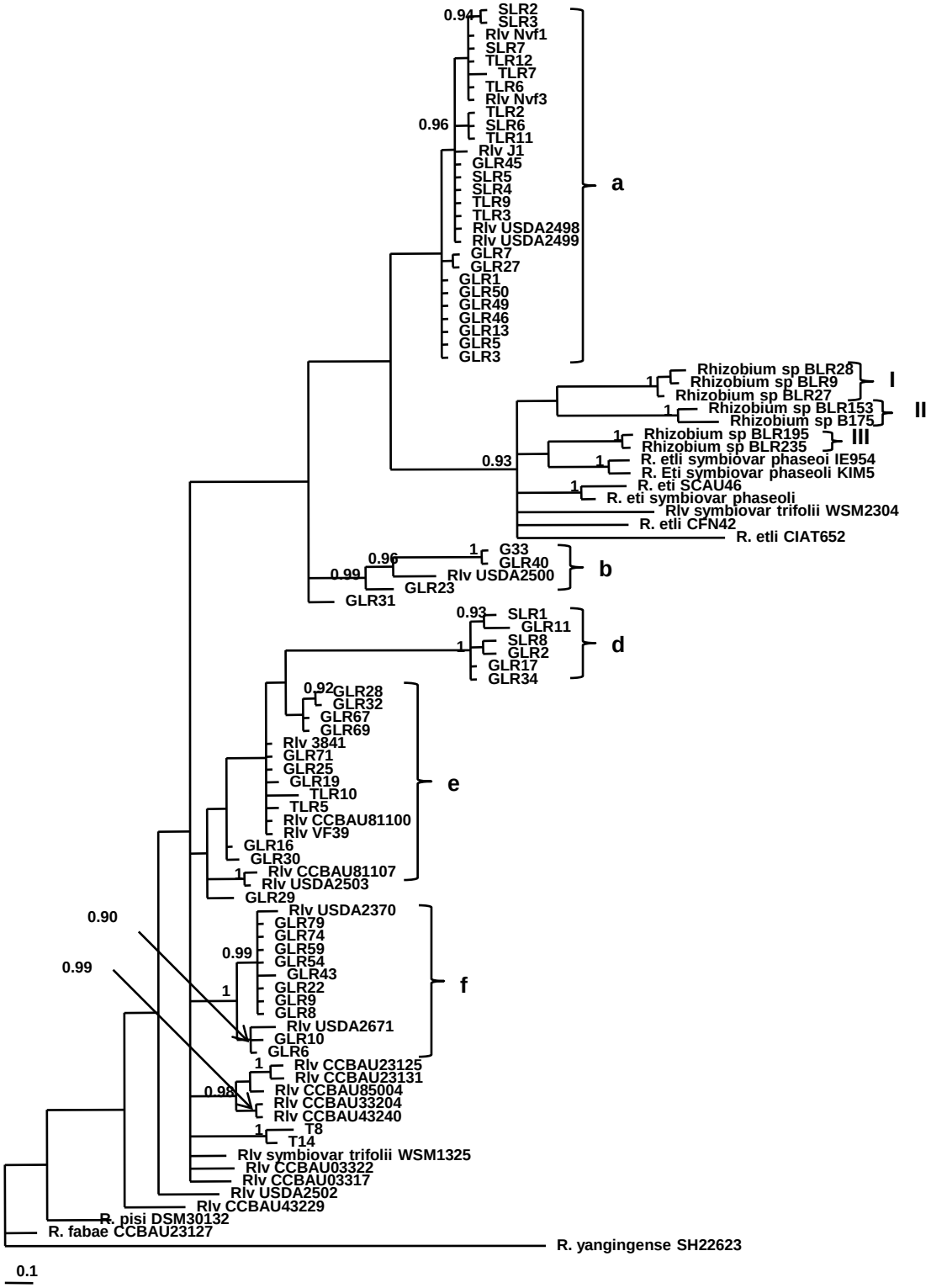
Appendix 10. Bayesian tree from partial sequences of 16S rRNA genes. Posterior probability values shown when ≥ 0.90. Abbreviations: GLR = German lentil rhizobia, TLR = Turkish lentil rhizobia, SLR = Syrian lentil rhizobia, R = *Rhizobium*, Rlv = *R. leguminosarum* symbiovar *viciae*.



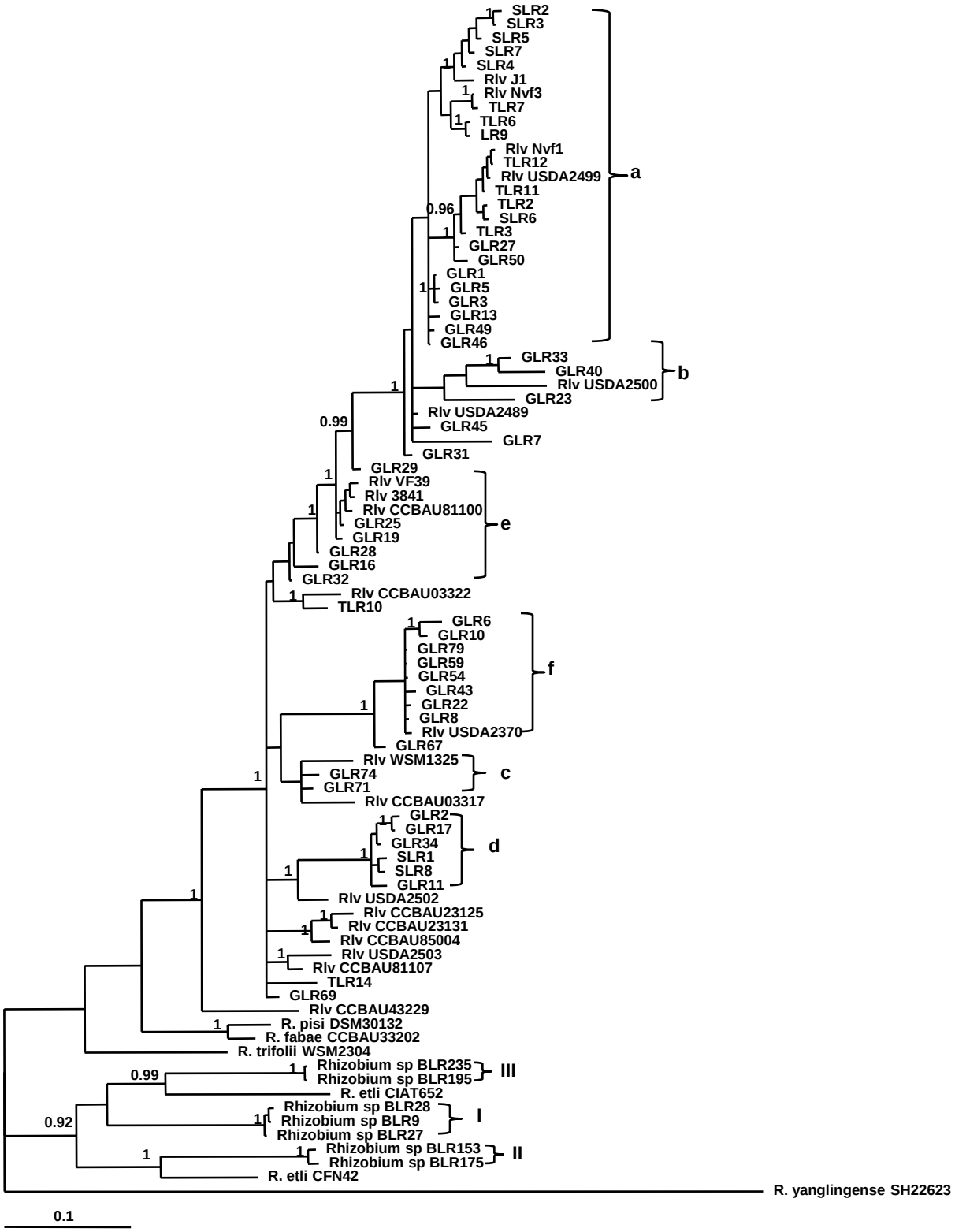
Appendix 11. Bayesian tree from partial sequences of *recA* gene. Posterior probability values shown when ≥ 0.90. Abbreviations: GLR = German lentil rhizobia, TLR = Turkish lentil rhizobia, SLR = Syrian lentil rhizobia, R = *Rhizobium*, Rlv = *R. leguminosarum* symbiovar *viciae*, I, II, III, IV = lineages , a – f = sub-lineages within lineage IV.



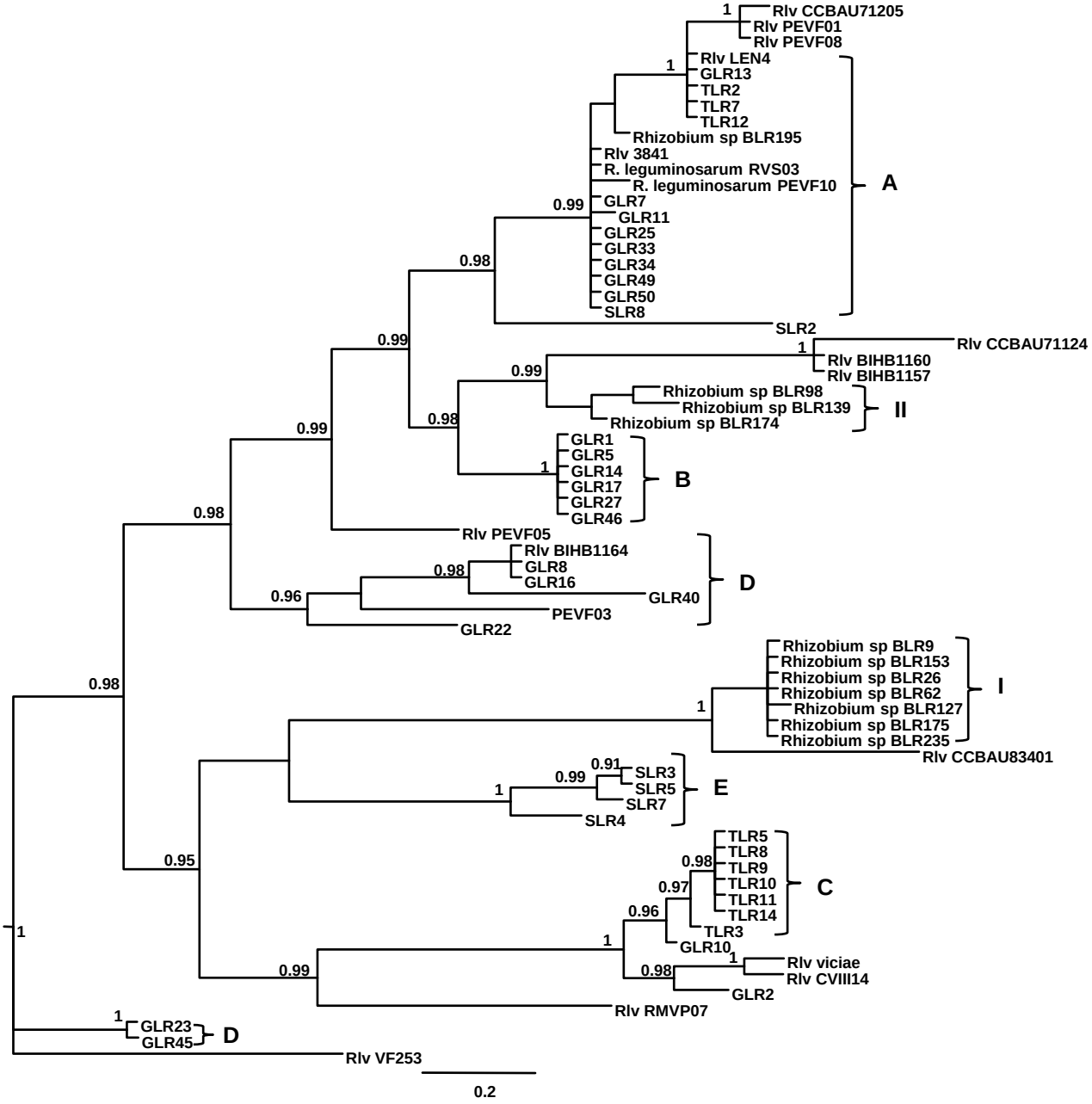
Appendix 12. Bayesian tree from partial sequences of *atpD* gene. Posterior probability values shown when ≥ 0.90. Abbreviations: GLR = German lentil rhizobia, TLR = Turkish lentil rhizobia, SLR = Syrian lentil rhizobia, R = *Rhizobium*, Rlv = *R. leguminosarum* symbiovar *viciae*, I, II, III, IV = lineages , a – f = sub-lineages within lineage IV.



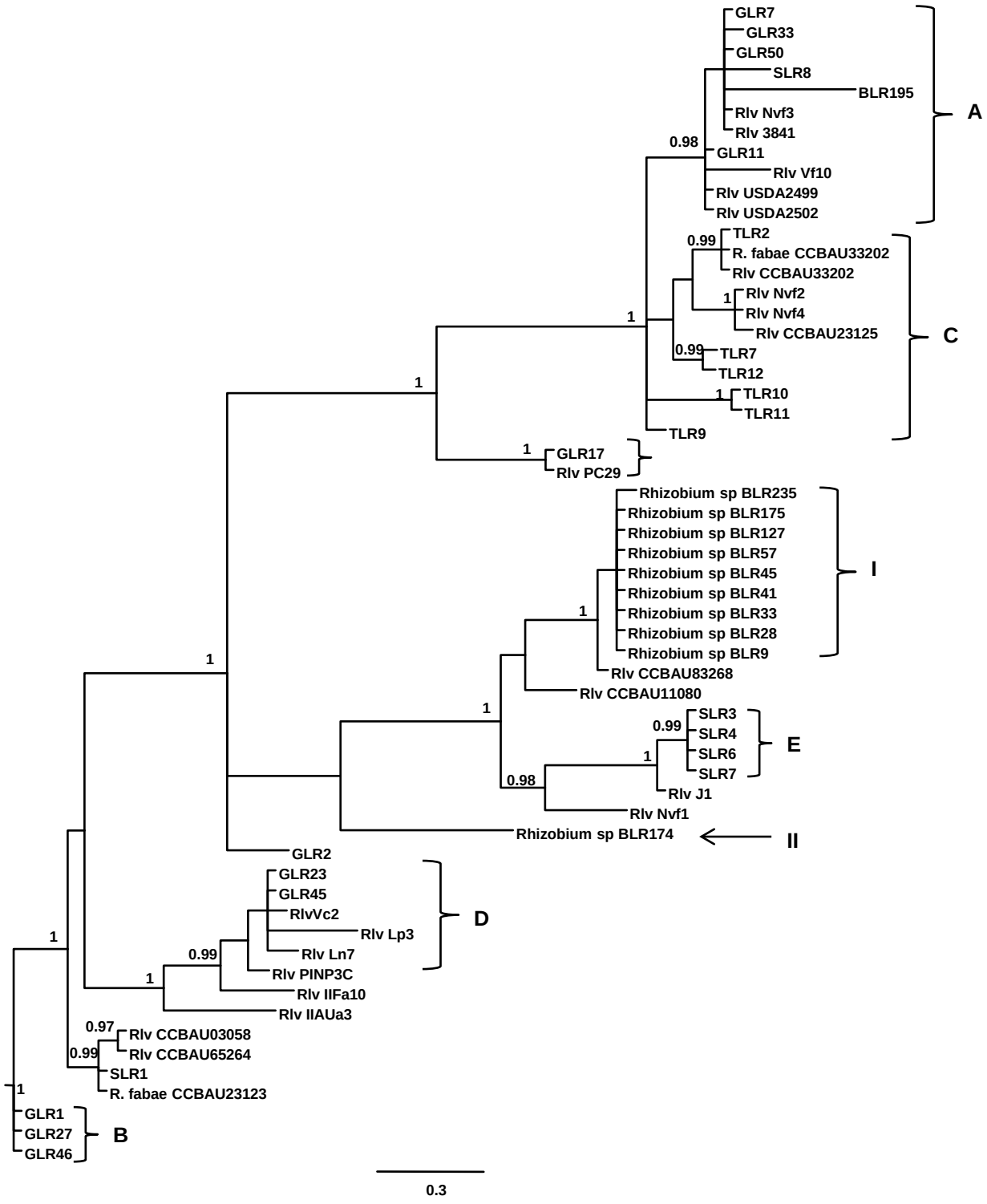
Appendix 13. Bayesian tree from partial sequences of *glnII* gene. Posterior probability values shown when ≥ 0.90. Abbreviations: GLR = German lentil rhizobia, TLR = Turkish lentil rhizobia, SLR = Syrian lentil rhizobia, R = *Rhizobium*, Rlv = *R. leguminosarum* symbiovar *viciae*, I, II, III, IV = lineages , a – f = sub-lineages within lineage IV.



Appendix 14. Bayesian tree from concatenated partial sequences of *recA-atpD-glnII* genes. Posterior probability values shown when ≥ 0.90. Abbreviations: GLR = German lentil rhizobia, TLR = Turkish lentil rhizobia, SLR = Syrian lentil rhizobia, R = *Rhizobium*, Rlv = *R. leguminosarum* symbiovar *viciae*, I, II, III, IV = lineages , a – f = sub-lineages within lineage IV.



Appendix 15. Bayesian tree from partial *nodC* gene sequences. Posterior probability values shown when ≥ 0.90. Abbreviation: GLR = German lentil rhizobia, TLR = Turkish lentil rhizobia, SLR = Syrian lentil rhizobia, BLR = Bangladeshi lentil rhizobia, Rlv = *R. leguminosarum* symbiovar *viciae*, Rl = *R. leguminosarum*, A – E = nodulation gene group from German, Turkish and Syrian isolates, I – II = nodulation gene group from Bangladeshi isolates.



Appendix 16. Bayesian tree from partial *nodD* gene sequences. Posterior probability values shown when ≥ 0.90. Abbreviation: GLR = German lentil rhizobia, TLR = Turkish lentil rhizobia, SLR = Syrian lentil rhizobia, BLR = Bangladeshi lentil rhizobia, Rlv = *R. leguminosarum* symbiovar *viciae*, Rl = *R. leguminosarum*, A – E = nodulation gene group from German, Turkish and Syrian isolates, I – II = nodulation gene group from Bangladeshi isolates.