Average nucleotide identity of genome sequences supports the description of Rhizobium lentis *sp. nov.,* Rhizobium bangladeshense *sp. nov. and* Rhizobium binae *sp. nov. from lentil (*Lens culinaris*) nodules.* M. Harun-or Rashid, J. Peter W. Young, Isobel Everall, Pia Clercx, Anne Willems, Markus Santhosh Braun and Michael Wink. International Journal of Systematic and Evolutionary Microbiology (2015) 65: 3037-3045.

DOI: 10.1099/ijs.0.000373

Title: Average nucleotide identity of genome sequences supports the description of *Rhizobium lentis* sp. nov., *Rhizobium bangladeshense* sp. nov. and *Rhizobium binae* sp. nov. from lentil (*Lens culinaris*) nodules

Running title: Rhizobia nodulating lentil

Content category: New Taxa Proteobacteria

Authors: M. Harun-or Rashid[1,4*]. J. Peter W. Young[2], Isobel Everall[2], Pia Clercx[3], Anne Willems[3], Markus Santhosh Braun[4] and Michael Wink[4]

Affiliation:
[1]Molecular Biology Laboratory, Soil Science Division, Bangladesh Institute of Nuclear agriculture (BINA), Bangladesh.

[2]Department of Biology, University of York, UK.

[3]Laboratory of Microbiology, Ghent University, Belgium.

[4]Department of Biology, Institute of Pharmacy and Molecular Biotechnology, Heidelberg University, Germany

Corresponding Author Present Address: **M. Harun-or Rashid**, Molecular Biology Laboratory, Soil Science Division, Bangladesh Institute of Nuclear agriculture (BINA), BAU Campus, Mymensingh-2202, Bangladesh.
Phone: +0088 01716 626 143; Fax: +0088 091 67842; e-mail: mhrashid08@gmail.com

Strain reference and accession number for type strain:

Rhizobium lentis: Type strain is BLR27^T (=LMG28441^T=DSMZ29286^T). Accession numbers for 16S rRNA: JN648905; *atpD*: JN648941; *recA*: JN649031; *glnII*: JN648976 and whole genome sequences: PRJEB7125 (http://www.ebi.ac.uk/ena/data/view/PRJEB7125).

Rhizobium bangladeshense: The type strain is BLR175^T (= LMG 28442^T = DSMZ 29287^T). Accession numbers for 16S rRNA: JN648931 ;*atpD*: JN648967; *recA*: JN649057; *glnII*: JN648979 and whole genome sequences:PRJEB7125 (http://www.ebi.ac.uk/ena/data/view/PRJEB7125).

Rhizobium binae : The type strain is BLR195^T(=LMG28443^T=DSM29288^T). Accession numbers for 16S rRNA: JN648932; *atpD*: JN648968; *recA*; JN649058; *glnII*: JN648980 and and whole genome sequences:PRJEB7125 (http://www.ebi.ac.uk/ena/data/view/PRJEB7125).

Abstract

Rhizobial strains isolated from effective root nodules of field-grown lentil (*Lens culinaris*) from different parts of Bangladesh were previously analyzed using four housekeeping genes (16S rRNA, *recA*, *atpD* and *glnII*) and three nodulation genes (*nodA*, *nodC* and *nodD*), DNA fingerprinting and phenotypic characterization. Analysis of housekeeping genes and DNA fingerprint indicated that the isolates belonged to three new clades in the genus *Rhizobium*. In the present study, a representative strain from each clade was further characterized by cellular fatty acid compositions, carbon substrate utilization pattern, DNA-DNA hybridization and average nucleotide identity (ANI) analyses from whole genome sequences. The DNA-DNA hybridization showed 50 – 56 % relatedness to their closest relatives (*Rhizobium etli* and *Rhizobium phaseoli*) and 50 – 60 % relatedness to each other. These results were further supported by average nucleotide identity values, based on genome sequencing, which were 87 – 92 % with their close relatives and 87 – 88 % with each other. On the basis of these results, three novel species, *Rhizobium lentis* sp. nov. (type strain BLR27^T), *Rhizobium bangladeshense* sp. nov. (type strain BLR175^T) and *Rhizobium binae* sp. nov. (type strain BLR195^T) are proposed. These species share common nodulation genes (*nodA*, *nodC* and *nodD*) that are similar to those of the symbiovar *viciae*.

Key words: Rhizobia, lentil, DNA-DNA relatedness, genome sequencing, average nucleotide identity

Rhizobia are nodule-forming nitrogen fixing bacteria that belong to the bacterial phylum Proteobacteria. Rhizobia can satisfy the nitrogen requirement of legumes by effective symbiosis with these plants and therefore are important bacteria for supporting plant growth and for environmental protection. About 180 species of nodule-forming bacteria in 12 different genera have been described. *Rhizobium* is an important genus of rhizobia and more than 80 species are currently included in this genus (http://www.bacterio.net/rhizobium.html). Among agricultural legume crops, lentil (*Lens culinaris*) is one of the oldest and remains very popular all over the world; it forms an effective symbiosis with rhizobia. *Rhizobium leguminosarum* symbiovar *viciae* (Rlv) is the main symbiont of the legume tribe Vicieae (Hou *et al*., 2009; Laguerre *et al*., 2003; Mutch & Young, 2004; Santillana *et al*., 2008; Tian *et al*., 2010 and many others), to which lentils belong, although other species, including *Rhizobium pisi*, *Rhizobium fabae* and *Rhizobium laguerreae* (Ramirez-Bahena *et al*., 2008; Saïdi *et al*., 2014; Tian *et al*., 2008), have also been described. Previous studies on rhizobia from lentil root nodules from different geographical locations revealed that Rlv is the major symbiont of lentil (Geniaux & Amarger, 1993; Hynes & O'Connell, 1990; Laguerre *et al*., 1992; Materon *et al*., 1995; Mowad & Beck, 1991). However, in our previous studies (Rashid *et al*., 2012 & 2014) we identified three new rhizobial clades / lineages in Bangladesh by multi locus sequence analysis, phenotypic and DNA fingerprint analyses. DNA-DNA hybridization (DDH) is an important experimental method to detect similarity between two genome sequences and a DDH value of 70% is the threshold value for bacterial species demarcation (McCarthy & Boltan, 1963). Wayne *et al.*, (1987) agreed that "the complete deoxyribonucleic acid (DNA) sequence would be the reference standard to determine phylogeny and that phylogeny should determine taxonomy", which was an impracticable goal at the time but is increasingly feasible (Chan *et al*., 2012). The approach known as Average Nucleotide Identity (ANI, Konstantinidis & Tiedje, 2005) provides relatedness information for prokaryotes at the whole genome level. ANI values equal to 95 – 96 %, calculated from pair-wise comparisons of shared sequences between two genomes, provides an equivalent value to DDH values of 70 %, which is the threshold value for bacterial species demarcation (Chan *et al*., 2012; Goris *et al*., 2007; Konstantinidis & Tiedje, 2005; Wayne *et al*., 1987).

Rhizobial strains were isolated from field-grown lentil root nodules from different parts of Bangladesh in 2009. Detailed descriptions of the strains (isolation localities, isolation procedure; identities; phylogenetic and population analysis; DNA fingerprint analysis and phenotypic characterization) are available in Rashid *et al*. (2012; 2014). DNA fingerprint analysis using ERIC-PCR, phylogeny and population analysis of *recA*, *atpD* and *glnII* genes strongly suggested that three clades were clearly distinct from all described *Rhizobium* species and belong to three novel species. In present study, we extended the phylogeny of the previously described 15 strains by including recently described rhizobial species. Three strains from these three lineages (BLR27^T, BLR175^T and BLR195^T from lineages I, II and III, respectively) were further characterized in this study by cellular fatty acid compositions, DNA-DNA hybridization and average nucleotide identity (ANI) to determine whether the proposal of new species was justified.

We described three new lineages in the genus *Rhizobium* from effective root nodules of lentils from Bangladesh in 2012 (Rashid *et al*., 2012). Since then, several additional species have been described in this genus. Therefore, in present study the 16S-rRNA, *recA* and *atpD* gene sequences from 15 strains from our previous study were compared with recently described species to assess their taxonomic status. The sequences were obtained from NCBI and aligned with the Clustal X (Thompson *et al*., 1997) in BioEdit (Hall, 1999). Phylogenetic trees were reconstructed using the neighbour-joining (NJ, Saitou & Nei, 1987) and maximum likelihood (ML, Rogers & Swofford, 1998) methods in MEGA version 5 (Tamura *et al*., 2011). For sequence evolution, general time reversible (GTR) model with gamma distribution was used in ML analysis. Bootstrap support for each node was evaluated with 1000 replicates. Trees were rooted using *Mesorhizobium* as outgroup. The 16S rRNA gene sequences from the strains from proposed new species (*R. lentis*, *R. bangladeshense* and *R. binae*) were very similar and clustered with type strains of *R. etli* CFN42^T, *R. phaseoli* CIAT652, *R. fabae* CCBAU33202^T, *R. pisi* DSM30132^T, *R. leguminosarum* USDA2370^T, *R. laguerreae* FB206^T (Fig. 1). The analysis of *recA*, *atpD* and *gln II* genes, and their concatenated sequences in our previous study (Rashid et al, 2012) and the concatenated sequences of *recA* and *atpD*

genes in Fig. 2 revealed that the proposed species formed three separate clades / lineages and the closest relatives were *R. etli*, and *R. phaseoli*. Moreover, *Rhizobium* sp. strain ESC1110, isolated from *Phaseolus vulgaris* from Hispaniola Island (Díaz-Alcántara *et al*., 2014) was closely related to *R. lentis* (Fig. 2).

Phylogenetic analysis of lentil rhizobia from Bangladesh in the present and previous studies showed that they belonged to the *Rhizobium leguminosarum* species complex, based on 16S rRNA gene sequences, but this gene did not provide a clear taxonomic identification since the sequences were >99% identical among these strains and with representatives of more than one closely related species. The phylogeny of the housekeeping genes *recA* and *atpD* was first used by Gaunt *et al*. (2001) and *glnII* gene by Turner & Young (2000); Stepkowski *et al*, (2005); and Vinuesa *et al*., (2005) to delineate the phylogeny of rhizobia and related bacteria with more confidence than was possible with 16S rRNA alone. These markers have been used widely and successfully in many studies of rhizobial diversity since then, although additional housekeeping genes may further improve the reliability of this approach (Martens *et al*., 2007, Vinuesa, 2010). Based on a phylogenetic analysis of these three housekeeping protein-coding genes, the lentil isolates fell into three clades that were distinct enough from each other and from known species to suggest that they might represent three new species. The phylogeny of housekeeping genes is particularly useful for placing multiple new isolates in relation to those described previously, as in this instance. It should be noted that, although the housekeeping gene phylogenies indicate that the lentil rhizobia from Bangladesh form three distinct lineages, their nodulation gene sequences do not reflect these lineages. Instead, the majority of strains, regardless of clade, share identical sequences for three genes involved in nodulation of the host plant, *nodA*, *nodC* and *nodD* (Rashid *et al*., 2012). These, and the variants found in the remaining strains, fall within the range of variation characteristic of symbiovar *viciae*. The symbiovar is a key attribute for the description of rhizobia: strains belonging to a symbiovar have similar nodulation genes and nodulate a similar range of hosts (Rogel *et al*., 2011). Importantly, strains that share the same symbiovar need not belong to the same species, because the nodulation genes are part of the accessory genome (Young *et al*.,

2006) and have frequently been subject to horizontal gene transfer within and between species (Young & Wexler, 1988). In this instance, all the lentil symbionts, regardless of species, belong to symbiovar *viciae* (Rashid *et al*., 2012).

Genetic diversity of the strains within each of the three clusters was assessed by high resolution ERIC-PCR, showing that the strains that belong to same cluster were not clonal (Rashid *et al*., 2012). For describing new rhizobial species, uncorrected genetic distances (similarity levels) correspond to an important parameter. Sequences of the *recA* and *atpD* genes of the three proposed species differed by 3.8 – 11.4% from those of the type strains of all other species in this clade of *Rhizobium* (Table S2). On the other hand, differences among strains within each species never exceeded 1.0% (Table S3), even though ERIC-PCR demonstrated that all the strains were genetically distinct (Rashid *et al*., 2012;Table S2 & S3).

High-quality DNA was prepared using the method of Wilson (Wilson, 1987) with minor modifications (Cleenwerck *et al*., 2002). DNA-DNA hybridizations were performed using a microplate method at 47.8 °C with photobiotin-labelled probes as described before (Goris *et al*., 2007) using an HTS7000 Bio Assay Reader (PE Applied Biosystems) for fluorescence measurements. The DNA G+C content was determined by HPLC as described previously (Mesbah *et al*., 1989). DDH experiments were conducted with the type strains of *R. etli* and *R. phaseoli*, since these two species were very close to the novel strains in phylogenetic analyses. The results of the DDH experiments are shown in Table 1. Strain BLR27^T (proposed type strain of clade I) showed 50 % and 56 % DNA relatedness to the type strains of *R. etli* and to *R. phaseoli,* respectively. It showed 60 % relatedness to strain BLR175^T (proposed type strain of clade II) and 50 % DNA relatedness to strain BLR195^T (proposed type strain of clade III). The strain BLR175^T showed 53 % relatedness to BLR195^T. Overall, the proposed type strains (BLR27^T, BLR175^T and BLR195^T) showed 50 – 62 % relatedness to their most closely related species and 50 – 60 % relatedness to each other. The DNA G+C content of the novel strains were 61.1 %, 60.9 % and 61.4 % for BLR27^T, BLR175^T and BLR195^T, respectively. These values are within the range of the genus *Rhizobium* (Jordan, 1984).

ANI is the best approach to determine genetic relatedness between two genomes because this method evaluates a large number of genes in its calculation, including slowly and fast evolving genes, and thus minimizes the effect of variable evolutionary rates or horizontal gene transfer events (Konstantinidis & Tiedje, 2005). Genomic DNA was extracted from strains grown in TY medium (Beringer, 1974) using PowerSoil DNA isolation kits (MoBio, Carlsbad, CA), and then fragmented, barcoded, quantitated and run as part of a batch of eight genomes on a 318 chip on an Ion Torrent PGM using the manufacturer's recommended protocols (Thermo Fisher, Waltham, MA). Each genome was assembled using the Newbler GS De Novo assembler version 2.8 (Roche diagnostics) with default parameter values. ANI was calculated within the JSpecies software (Richter & Rosselló-Móra, 2009). The Nucleotide MUMmer algorithm (NUCmer) was used, with default parameter settings, to calculate the ANI by subtracting the similarity errors from the alignment length (Kurtz *et al*., 2004; Richter & Rosselló-Móra, 2009). Genomes were compared with each other, with genome assemblies obtained using the same methods for *Rhizobium pisi* DSM30132^T and *Rhizobium fabae* CCBAU33202^T (unpublished), and with complete genome assemblies downloaded from NCBI for the following strains: *Rhizobium etli* CFN42^T (GCA_000092045), *Rhizobium phaseoli* CIAT652 (GCA_000020265), *Rhizobium leguminosarum*3841 (GCA_000009265), *Rhizobium leguminosarum* WSM1325 (GCA_000023185), *Rhizobium* sp. WSM2304 (GCA_000021345). Ion Torrent sequencing yielded 155 Mbp, 264 Mbp, 237 Mbp of sequence from BLR27^T, BLR175^T and BLR195^T, respectively, corresponding to 27- to 49-fold coverage, so it can be expected that virtually all the genomic sequence is included. Assembly resulted in 140, 89 and 187 contigs (>100 bp) with N50 sizes of 229 kb, 286 kb and 173 kb, respectively. Pairwise ANI was calculated between these genomic sequences, and with other strains in the *R. leguminosarum* species complex for which genome data were available (Table 2). Each ANI was calculated in both directions, but the results never differed by more than 0.03%. ANI values ranged from 87.27% to 92.39%, with this highest value being between BLR195^T and *R. phaseoli* CIAT652. All these values are well below 96%, which is an accepted value as the species boundary, equivalent to a DDH of 70% (Richter & Rosselló-Móra,

2009). Hence, each of the three BLR strains belongs to a distinct species, and these are different from all the closely related species described previously.

Different phenotypic characteristics i.e. colony size, growth on LB medium, acid-alkali production, tolerance to NaCl, temperature and pH, and antibiotic sensitivity were determined following the protocols described in Rashid *et al*., (2012). To observe the host range of proposed species for nodule formation, cross inoculation tests were performed with pea (*Pisum sativum*) and lathyrus (*Lathyrus sativa*) (Rashid *et al*., 2012). Randomly selected all strains from three species were availed to form nodules with both pea and lathyrus suggesting that proposed new species are ideal members of symbiovar *viciae*. Utilization of different carbon substrates by proposed species and close relatives were determined using Biolog GENIII following manufacturer's instructions and results are mentioned in Table S1.The cellular fatty acid compositions of type strain of proposed species were analyzed after growing on YEMA plates at 28 ^{o}C for 3 days. Cells were saponified and transmethylated as described by Kuykendal *et al.* (1988) and were separated by using the Sherlock microbial identification system (RTSBA6; MIDI) and an Agilent (model 680N) gas chromatograph, and were determined at DSMZ, Germany. The results of fatty acid analysis are mentioned in Table S4. Phenotypic characteristics of strains belonging to the three proposed species and their close relatives are given in Table S5. *R. binae* and most *R. bangladeshense* strains were able to grow at pH 10, and also in 0.5 % NaCl, unlike most strains of *R. lentis*. On the other hand, all *R. lentis* and some *R. bangladeshense* strains showed resistance to ampicillin compared to their close relative (*R. etli*). The type strains of the proposed species and their close relatives shared few common fatty acids: 16:0; 18:0; 16:0 3 OH; 19:0 cyclo ω8c; summed feature 2 and summed feature 8 but the amounts were different with close relatives and among themselves. Moreover, 15:0 iso 2 OH and summed feature 3 are found in two of the proposed species. Carbon and nitrogen substrates utilization pattern of proposed novel species differed with their close relatives and among themselves.

We have selected one representative strain from each clade that was identified in the housekeeping gene phylogeny, and demonstrate through the use of DDH and ANI analysis that they do indeed meet the standard criteria for distinct species, since all DDH values are below 70% and ANI values below 96%. DDH has been the standard method for bacterial species demarcation for the last 50 years (McCarthy & Bolton,1963; Tindall *et al*., 2010, Wayne *et al*., 1987), but it has major limitations. It is time-consuming, laborious, and hard to standardize between laboratories. An increasingly significant limitation, as more species are described, is that DDH requires a laboratory comparison with all possible close relatives. DDH was developed before genome sequencing became feasible, but sequence-based methods have the potential to provide more reliable information more easily. The calculation of average nucleotide identity (ANI) from genome sequence data has been shown to give comparable results to DDH, with a species boundary at around 96% (Goris *et al*., 2007; Konstantindis & Tiedje, 2005). Our study does not test the accuracy of this boundary, since all ANI values were much lower than this, but it does provide further evidence that low ANI values can be used as an effective substitute for DDH when establishing that strains do not belong to the same species. We expect that, in future, proposals for new bacterial species will increasingly use ANI evidence in place of DDH. In this study, the two methods provide consistent evidence that three new species are involved in lentil nodulation in Bangladesh, for which we proposed the names *Rhizobium lentis*, *Rhizobium bangladeshense* and *Rhizobium binae*.

Description of *Rhizobium lentis*

Rhizobium lentis (len'tis. L. gen. n. lentis, referring to *Lens*, the plant genus from which the bacteria were isolated).

Cells are Gram negative, aerobic, non-spore forming and rod shaped. The optimum growth temperature is 28 °C at pH 7. Colonies are circular, convex and creamy white on YEMA medium. Strains grow at 12 – 37 °C but can survive at 4 °C. Strains grow well at pH 5.5 to 8.2 and are sensitive to 0.5 % NaCl in YEMA medium. Most of the strains are resistant to ampicillin, kanamycin and nalidixic acid. Strains do not tolerate tetracycline and do not show any growth on LB medium. The fatty acid composition of type strain is

15:0 iso 2-OH, 16:0, 16:0 3-OH, 18:0, 19:0 cyclo ω8c, summed featured 2, summed featured 3 and summed featured 8. In Biolog III systems, type strain could utilize α-D lactose, β-methyl-D-glucoside, D-sorbitol, D-mannito,D-arbitol-glycerol, D-fructose-6-phosphate, L-aspertic acid, D-gluconic acid, mucic acid, D-lactic acid methyl ester, L-lactic acid, L-histidine, L-lactic acid, β-hydroxy-D, L-butyric acid, D-malic acid, L-malic acid, acetic acid and formic acid. Type strain was unable to utilize D-maltose, D-trehalose, D-cellobiose, gentiobiose, sucrose, D-raffinose, α-D-glucose, D-turanose, D-melibose, mannose, galactose, 3-methyle glucose, inosine, D-aspertic acid, glycyl-L-proline, L-alanine, L-aspertic acid, L-arginine, L-serine, pectine, D-saccharic acid, p-hydroxy-phenylacetic acid, methyl pyruvate, citric acid, bromo-succinic acid, acetoacetic acid or propionic acid. Type strain could grow in the presence of lincomycin, tetrazolium violet, tetrazolium blue and nalidixic acid, but not with 1% sodium lactate, troleandomycin, lithium chloride,potassium tellurite or sodium butyrate.

Type strain is BLR27^T (= LMG 28441^T = DSMZ 29286^T). The DNA G+C content of type strain is 61.1%. Type strain was isolated from effective nodules of *Lens culinaris* from Bagatipara, Natore district of Bangladesh. Other strains (BLR9, BLR26, BLR28, BLR29, BLR33, BLR41, BLR45, BLR59, BLR87, BLR98, BLR100, BLR122, BLR127, BLR137, BLR139 and BLR160) were isolated from different parts of Bangladesh.

Description of *Rhizobium bangladeshense*

Rhizobium bangladeshense, (ban.gla.desh.en'se. N.L. neut. adj. *bangladeshense*, from Bangladesh, referring to the geographical origin of the strains).

Cells are Gram negative and rod shaped. Colonies are circular, convex and creamy white on YEMA medium. The optimum temperature for growth is 28 °C at pH 7, although strains could grow well up to 37 °C. Strains can tolerate pH values between 5.5 and 10 and are sensitive to ampicillin, resistant to kanamycin and nalidixic acid, and grow well in YEMA medium containing 0.5 % NaCl. Strains do not tolerate tetracycline and do not show any growth on LB medium. Fatty acid composition of type strain is 16:0, 16:0 3-OH, 18:0, 19:0 cyclo ω8c, summed featured 2 and summed featured 8. In Biolog III systems type strain could utilize D-maltose, D-trehalose, D-cellobiose, gentiobiose, sucrose, D-raffinose, α-D-glucose, D-turanose, α-D lactose, D-fructose, β-

methyl-D-glucoside, salicin, N-acetyl-D-galactosamine, D-sorbitol, D-mannitol, D-arbitol, glycerol, D-glucose-6-phosphate, D-gluconic acid, quinic acid, D-saccharic acid, D-lactic acid methyl ester, lactic acid, α-keto-glutaric acid and tween 40. Type strain failed to utilize dextrin, D-aspertic acid, glycyl-L-proline, L-alanine, L-arginine, L-glutamic acid, L-histidine, L-serine, mucic acid, p-hydroxy-phenylacetic acid, methyl pyruvate, citric acid, D-malic acid, L-malic acid, propionic acid or formic acid. Type strain could grow in the presence of lincomycin and potassium tellurite, but not with 1% sodium lactate, troleandomycin, tetrazolium violet, tetrazolium blue, Nalidixic acid, lithium chloride and sodium butyrate.

Type strain is BLR175^T (= LMG 28442^T = DSMZ 29287^T). The DNA G+C content of type strain is 61 %.Type strain was isolated from effective nodules of *Lens culinaris* from Mohammedpur, Khulna district of Bangladesh. Other strains (BLR62, BLR99, BLR129, BLR153 and BLR154) were isolated from different parts of Bangladesh.

Description of *Rhizobium binae*

Rhizobium binae (bi'nae. N.L. gen. fem. n. binae, abbreviation for Bangladesh Institute of Nuclear Agriculture, a research institute where the first steps to isolate the bacteria were taken).

Cells are Gram negative and rod shaped. Colonies are circular, convex and creamy white on YEMA medium. The optimum temperature for growth is 28 °C at pH 7, but the strains grow well at 37 °C. Strains survive at pH values between 5.5 and 10 and tolerate 1 % NaCl in YEMA. They are very sensitive to ampicillin and resistant to kanamycin and nalidixic acid. Strains do not tolerate tetracycline and do not show any growth on LB medium. Fatty acid composition of type strain is 15:0 iso 2-OH, 16:0, 16:0 3-OH, 18:0, 18:1ω9c, 18:0 3-OH, 18:1ω7c 11methyl, 19:0 cyclo ω8c, summed feature 2 and summed feature 8. Type strain could utilize dextrin, D-maltose, D-trehalose, D-cellobiose, gentiobiose, sucrose, D-raffinose, α-D-glucose, D-turanose, α-D lactose, D-fructose, D-melibiose, β-methyl-D-glucoside, salicin, N-acetyl-D-galactosamine, D-mannose, D-galactose, D-mannitol, D-sorrbitol, D-arabitol, glycerol, D-glucose-6-phosphate, D-fructose-6-phosphate, D-alanine, L-aspartic acid, L-histidine, l-pyroglutamic acid, quinic acid, D-saccharic acid, methyl pyruvate, L-lactic acid, citric acid, D-malic acid, L-malic acid, bromo-succinic acid, β-hydroxy-d,l-butyric acid and

acetic acid. Type strain failed to utilize N-acetyle-D-mannosamine, 3-methyle glucose, inosine, glycyl-L-proline, L-arginine, D-galacturonic acid, D-glucuronic acid, glucuronamide, p-hydroxy-phenylacetic acid, D-lactic acid methyl ester, α-keto-glutaric acid, tween 40, propionic acid or formic acid. Type strain could grow in the presence of lincomycin and potassium tellurite but not with 1% sodium lactate, troleandomycin, lithium chloride or sodium butyrate.

Type strain is BLR195^T (=LMG 28443^T = DSM 29288^T). The DNA G+C content of type strain is 61.5 %. Type strain was isolated from effective nodules of *Lens culinaris* from Sarishadi, Feni district of Bangladesh. All strains (BLR 228, BLR235 and the type strain) were isolated from the Southeast part of Bangladesh.

Acknowledgements

This work was supported by the European Union's Seventh Framework Programme (BRIO project, grant KBBE 2010-4-266106 and LEGATO project, grant KBBE 2013-7-613551) and was also supported by BINA Bangladesh. We thank David Sherlock and Sally James (University of York) for preparing and sequencing the genomic DNA. We are grateful to Dr. Martin Krehenbrink (Cysal Germany) and Prof. M.A. Hashem (Bangladesh Agricultural University) for reading our manuscript.

References

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

Chan, J. Z-M., Halachev, M.R., Loman, N.J., Constantinidou, C.& Pallen, K.J. (2012). Defining bacterial species in the genomic era: insight from the genus *Acinetobacter*. *BMC Microbiol* **12**, 302.

Cleenwerck, I., Vandemeulebroecke, K., Janssens, D. & Swings, J. (2002). Re-examination of the genus *Acetobacter*, with descriptions of *Acetobacter cerevisiae* sp. nov. and *Acetobacter malorum* sp. nov. *Int J Syst Evol Microbiol* **52**, 1551 – 1558.

Díaz-Alcántara, C-A., Ramírez-Bahena, M-H., Mulas, D., García-Fraile, P., Gómez-Moriano, A., Peix, A., Velázquez, E. & González-Andrés, F. (2014). Analysis of rhizobial strains nodulating *Phaseolus vulgaris* from Hispaniola Island, a geographic bridge between Meso and South America and the first historical link with Europe. *Syst Appl Microbiol* **37**, 149 – 156.

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

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

Goris, J., Konstantinidis, K.T., Klappenbach, J.A., Coenye, T., Vandamme, P. & Tiedje, J.M. (2007).. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol* **57**, 81 – 91.

Hall, T.A. (1999). Bio-Edit: a user friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic acids symp ser* **41**, 95 – 98.

Hou, B.C., Wang, E.T., Li, Y., Jia, R.Z., Chen, W.F., Man, C.X., Sui, X.H. & Chen, W.X. (2009). Rhizobial resource associated with epidemic legumes in Tibet. *Microb Ecol* **57**, 69 – 81.

Hynes, M.F. & O'Connell, M.P. (1990). Host plant effect on competition among strains of *Rhizobium leguminosarum*. *Can.J Microbiol* **36**, 864 – 869.

Jordan, D.C. (1984). Family III. Rhizobiaceae. In *Bergey's Manual of Systematic Bacteriology.* pp. 234 – 242. Edited by Krieg, N.R. & Holt, J.G, Williams and Wilkins, Baltimore, U.S.A.

Konstantinidis, K.T. & Tiedje, J.M. (2005). Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci* U.S.A. **102**, 2567 – 2572.

Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C. & Salzberg, S.L. (2004). Versatile and open software for comparing large genomes. *Genome Biol* **5**, 12.

Kuykendall, L.D., Roy, M.A., O'Neill, J.J. & Devine, T.E. (1988) Fatty acids, antibiotic resistance and deoxyribonucleic acid homology groups of *Bradyrhizobium japonicum*. *Int J Syst Bacteriol* **38**, 358 – 361.

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

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

Martens, M., Delaere, M., Coopman, R., De Vos, P., Gillis, M. & Willems, A. (2007). Multilocus sequence analysis of *Ensifer* and related taxa. *Int J Syst Evol Microbiol* **57**, 489 – 503.

Materon, L.A., Keatinge, J.D.H., Beck, D.P., Yurtsever, N., Karuc, K. & Altuntas, S. (1995). The role of rhizobial biodiversity in legume crop productivity in the West Asian highlands. *Expt Agri* **31**, 485 – 491.

McCarthy, B.J. & Bolton, E.T. (1963). An approach to the measurement of genetic relatedness among organisms. *Proc Natl Acad Sci* U.S.A. **50**, 156 – 164.

Mesbah, M., Premachandran, U. & Whitman, W.B. (1989). Precise measurement of the G+C content of deoxyribonucleic acid by high-performance liquid chromatography. *Int J Syst Bacteriol* **39**, 159 – 167.

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

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

Rashid, M.H., Gonzalez, H., Young, J.P.W. & Wink, M. (2014). *Rhizobium leguminosarum* is the symbiont of lentil in the Middle East and Europe but not in Bangladesh. *FEMS Microbiol Ecol* **87**, 64 – 77.

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

Ramírez-Bahena, M.H., García-Fraile, P., Peix, A., Valverde, A., Rivas, R., Igual, J.M., Mateos, P.F., 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.

Richter, M. & Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci* U.S.A. **106**, 19126 – 19131.

Rogel, M.A., Ormeño-Orrillo, E. & Martínez Romero, E. (2011). Symbiovars in rhizobia reflect bacterial adaptation to legumes. *Syst Appl Microbiol* **34**, 96 – 104.

Rogers, J.S. & Swofford, D.L. (1998). A fast method for approximating maximum likelihoods of phylogenetic trees from nucleotide sequences. *Syst Biol* **47**, 77 – 89.

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

Santillana, N., Ramírez-Bahena, M.H., García-Fraile, P., Velázquez, E. & Zúñiga, 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.

Segovia, L., Young, J.P.W. & 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.

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

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.

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

Tian, C.F., Young, J.P.W., Wang, E.T., Tamimi, S.M. & Chen, W.X. (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, C.F., Wang, E.T., Wu, L.J., Han, T.X., Chen, W.F., Gu, C.T., Gu, J.G. & Chen, W.X. (2008). *Rhizobium fabae* sp.nov., a bacterium that nodulates *Vicia faba*. *Int J Syst Evol Microbiol* **58**, 2871 – 2875.

Tighe, S. W., de Lajudie, P., Dipietro, K., Lindström, K., Nick, G. & Jarvis, B.D.W. (2000). Analysis of cellular fatty acids and phenotypic relationships of *Agrobacterium*, *Bradyrhizobium*, *Mesorhizobium*, *Rhizobium* and *Sinorhizobium* species using the Sherlock Microbial Identification System. *Int J Syst Evol Microbiol* **50**, 787 – 801.

Tindall, B.J., Rosselló-Móra, R., Busse, H.J., Ludwig, W. & Kämpfer, P. (2010). Notes on the characterization of prokaryote strains for taxonomic purposes. *Int J Syst Evol Microbiol* **60**, 249 – 266.

509 **Turner, S. L., & Young, J. P. W. (2000).** The glutamine synthetases of rhizobia: phylogenetics
510 and evolutionary implications. *Mol Biol Evol* **17,** 309 – 319.
511
512 **Vinuesa, P. (2010)**. Multilocus sequence analysis and species phylogeny. In: Molecular
513 Phylogeny of Microorganisms, Oren, A., Papke, R.T. (Eds). Caister Academic Press, Norfolk,
514 UK. pp.41 – 64.
515
516 **Vinuesa, P., Silva, C., Werner, D. & Martínez-Romero, E. (2005).** Population genetics and
517 phylogenetic inference in bacterial molecular systematics: the roles of migration and
518 recombination in Bradyrhizobium species cohesion and delineation. *Mol Phylogenet Evol* **34,** 29
519 – 54.
520
521 **Wayne, L.G., Brenner, D.G., Colwell, R.R., Grimont, P.A.D., Kandler, O., Krichevsky, M.I.,**
522 **Moore, L.H., Moore, W.E.C., Murray, R.G.E.** & **other authors (1987).** Report of the Ad Hoc
523 committee on reconciliation of approaches to bacterial systematic. *Int J Syst Bacteriol* **37**, 463 –
524 464.
525
526 **Wilson, K. (1987).** Preparation of genomic DNA from bacteria. In *Current Protocols in Molecular*
527 *Biology*, pp. 241 – 245. Edited by Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D.,
528 Seidman, J.G., Smith, J.A., Struhl, K.. Green Publishing and Wiley-Inter Science.
529
530 **Young, J.P.W., Crossman, L.C., Johnston, A.W.B., Thomson, N.R., Ghazoui, Z.F., Hull**,
531 K.H., **Wexler, M., Curson, A.R.J., Todd, J.D. & other authors (2006).** The genome of
532 *Rhizobium leguminosarum* has recognizable core and accessory components. *Genome Biol* **7**,
533 R34.
534
535 **Young, J.P.W. & Wexler, M. (1988).** Sym plasmid and chromosomal genotypes are correlated
536 in field populations of *Rhizobium leguminosarum*. *J Gen Microbiol* **134**, 2731 – 2739.
537
538
539
540
541

Fig. 1. ML tree from 16S rRNA gene partial sequences. Bootstrap values ≥70 are indicated for each node (1000 replicates).

559 **Fig. 2**. ML tree from concatenated sequence of atpD-recA genes. Bootstrap
560 values ≥70 are indicated for each node (1000 replicates).

561 **Tables**

562 **Table 1.** Genetic relatedness measured by DDH between strains representing the novel clades and type strains of the
563 most closely related species

Species	Strain	*DNA G+C content (% mol)*	DNA-DNA relatedness (%)				
			Rhizobium lentis (BLR27^T)	*Rhizobium bangladeshense* (BLR175^T)	*Rhizobium binae* (BLR195^T)	*Rhizobium etli* (LMG 17827^T)	*Rhizobium phaseoli* (LMG 8819^T)
Rhizobium lentis	BLR 27^T	61.1	100.0	60.9	50.4	50.6	56.3
Rhizobium bangladeshense	BLR175^T	61.0	60.9	100.0	53.2	49.2	62.2
Rhizobium binae	BLR195^T	61.5	50.4	53.2	100.0	54.9	55.8

564

565

566

567

568 **Table 2.** Average nucleotide identity (ANI) between strains representing the novel clades and the most closely related
569 sequenced members of the *Rhizobium leguminosarum* complex

Species	Strain	Average Nucleotide Identity (%)		
		R. lentis BLR 27^T	*R. bangladeshense* BLR 175^T	*R. binae* BLR 195^T
R. lentis	BLR27^T	100	89.70	88.64
R. bangladeshense	BLR175^T	89.72	100	88.51
R. binae	BLR195^T	88.62	88.51	100
R. etli	CFN 42^T	89.04	88.64	89.28
R. phaseoli	CIAT 652	88.56	88.25	92.39
R. fabae	CCBAU 33202^T	88.90	88.64	88.81
R. pisi	DSM 30132^T	88.82	88.18	88.46
R. leguminosarum	3841	88.33	87.27	88.28
R. leguminosarum	WSM1325	88.01	87.59	88.06
Rhizobium sp.	WSM2304	88.09	87.76	88.55

570 Abbreviation: *R.* = *Rhizobium*

571

572
573

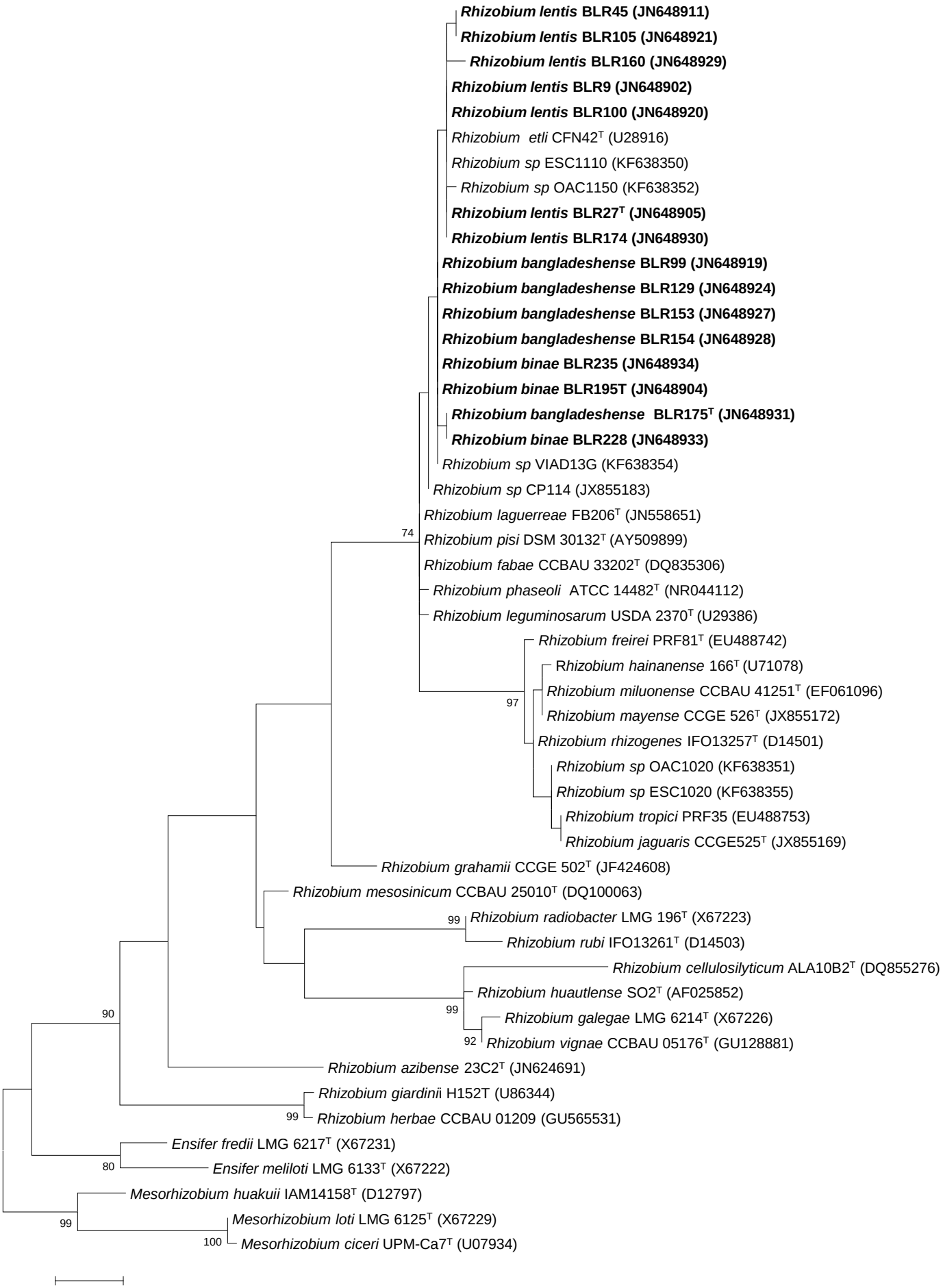
Rhizobium lentis BLR45 (JN648911)
Rhizobium lentis BLR105 (JN648921)
Rhizobium lentis BLR160 (JN648929)
Rhizobium lentis BLR9 (JN648902)
Rhizobium lentis BLR100 (JN648920)
Rhizobium etli CFN42^T (U28916)
Rhizobium sp ESC1110 (KF638350)
Rhizobium sp OAC1150 (KF638352)
Rhizobium lentis BLR27^T (JN648905)
Rhizobium lentis BLR174 (JN648930)
Rhizobium bangladeshense BLR99 (JN648919)
Rhizobium bangladeshense BLR129 (JN648924)
Rhizobium bangladeshense BLR153 (JN648927)
Rhizobium bangladeshense BLR154 (JN648928)
Rhizobium binae BLR235 (JN648934)
Rhizobium binae BLR195T (JN648904)
Rhizobium bangladeshense BLR175^T (JN648931)
Rhizobium binae BLR228 (JN648933)
Rhizobium sp VIAD13G (KF638354)
Rhizobium sp CP114 (JX855183)
Rhizobium laguerreae FB206^T (JN558651)
74
Rhizobium pisi DSM 30132^T (AY509899)
Rhizobium fabae CCBAU 33202^T (DQ835306)
Rhizobium phaseoli ATCC 14482^T (NR044112)
Rhizobium leguminosarum USDA 2370^T (U29386)
Rhizobium freirei PRF81^T (EU488742)
Rhizobium hainanense 166^T (U71078)
Rhizobium miluonense CCBAU 41251^T (EF061096)
97
Rhizobium mayense CCGE 526^T (JX855172)
Rhizobium rhizogenes IFO13257^T (D14501)
Rhizobium sp OAC1020 (KF638351)
Rhizobium sp ESC1020 (KF638355)
Rhizobium tropici PRF35 (EU488753)
Rhizobium jaguaris CCGE525^T (JX855169)
Rhizobium grahamii CCGE 502^T (JF424608)
Rhizobium mesosinicum CCBAU 25010^T (DQ100063)
99
Rhizobium radiobacter LMG 196^T (X67223)
Rhizobium rubi IFO13261^T (D14503)
Rhizobium cellulosilyticum ALA10B2^T (DQ855276)
Rhizobium huautlense SO2^T (AF025852)
99
90
Rhizobium galegae LMG 6214^T (X67226)
92
Rhizobium vignae CCBAU 05176^T (GU128881)
Rhizobium azibense 23C2^T (JN624691)
Rhizobium giardinii H152T (U86344)
99
Rhizobium herbae CCBAU 01209 (GU565531)
Ensifer fredii LMG 6217^T (X67231)
80
Ensifer meliloti LMG 6133^T (X67222)
Mesorhizobium huakuii IAM14158^T (D12797)
99
Mesorhizobium loti LMG 6125^T (X67229)
100
Mesorhizobium ciceri UPM-Ca7^T (U07934)
0.01

Fig. 1.

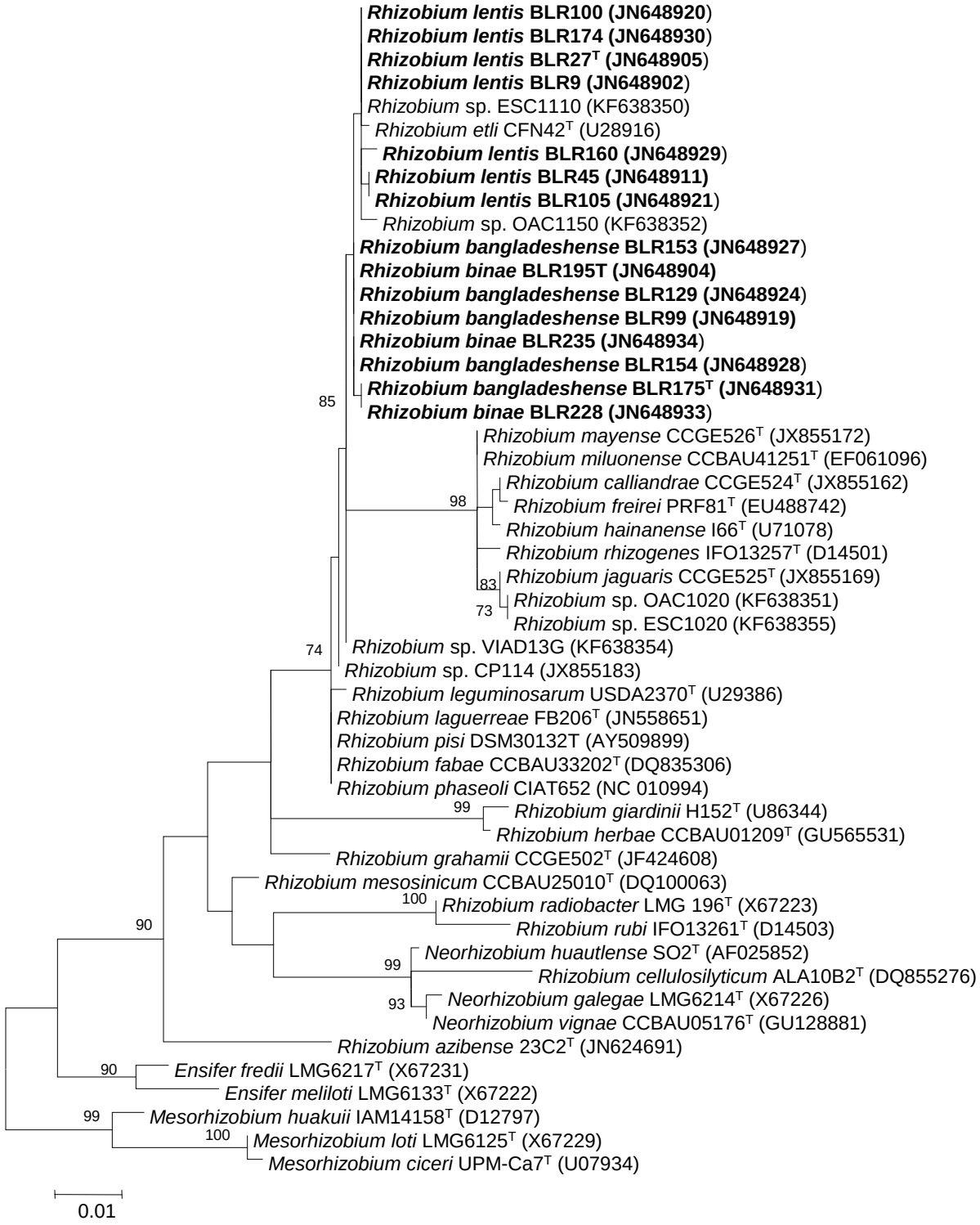
Rhizobium lentis BLR100 (JN648920)
Rhizobium lentis BLR174 (JN648930)
Rhizobium lentis BLR27T (JN648905)
Rhizobium lentis BLR9 (JN648902)
Rhizobium sp. ESC1110 (KF638350)
Rhizobium etli CFN42T (U28916)
Rhizobium lentis BLR160 (JN648929)
Rhizobium lentis BLR45 (JN648911)
Rhizobium lentis BLR105 (JN648921)
Rhizobium sp. OAC1150 (KF638352)
Rhizobium bangladeshense BLR153 (JN648927)
Rhizobium binae BLR195T (JN648904)
Rhizobium bangladeshense BLR129 (JN648924)
Rhizobium bangladeshense BLR99 (JN648919)
Rhizobium binae BLR235 (JN648934)
Rhizobium bangladeshense BLR154 (JN648928)
Rhizobium bangladeshense BLR175T (JN648931)
Rhizobium binae BLR228 (JN648933)
Rhizobium mayense CCGE526T (JX855172)
Rhizobium miluonense CCBAU41251T (EF061096)
Rhizobium calliandrae CCGE524T (JX855162)
Rhizobium freirei PRF81T (EU488742)
Rhizobium hainanense I66T (U71078)
Rhizobium rhizogenes IFO13257T (D14501)
Rhizobium jaguaris CCGE525T (JX855169)
Rhizobium sp. OAC1020 (KF638351)
Rhizobium sp. ESC1020 (KF638355)
Rhizobium sp. VIAD13G (KF638354)
Rhizobium sp. CP114 (JX855183)
Rhizobium leguminosarum USDA2370T (U29386)
Rhizobium laguerreae FB206T (JN558651)
Rhizobium pisi DSM30132T (AY509899)
Rhizobium fabae CCBAU33202T (DQ835306)
Rhizobium phaseoli CIAT652 (NC 010994)
Rhizobium giardinii H152T (U86344)
Rhizobium herbae CCBAU01209T (GU565531)
Rhizobium grahamii CCGE502T (JF424608)
Rhizobium mesosinicum CCBAU25010T (DQ100063)
Rhizobium radiobacter LMG 196T (X67223)
Rhizobium rubi IFO13261T (D14503)
Neorhizobium huautlense SO2T (AF025852)
Rhizobium cellulosilyticum ALA10B2T (DQ855276)
Neorhizobium galegae LMG6214T (X67226)
Neorhizobium vignae CCBAU05176T (GU128881)
Rhizobium azibense 23C2T (JN624691)
Ensifer fredii LMG6217T (X67231)
Ensifer meliloti LMG6133T (X67222)
Mesorhizobium huakuii IAM14158T (D12797)
Mesorhizobium loti LMG6125T (X67229)
Mesorhizobium ciceri UPM-Ca7T (U07934)
85
98
83
73
74
99
100
90
99
93
90
99
100
0.01

Fig. 2.

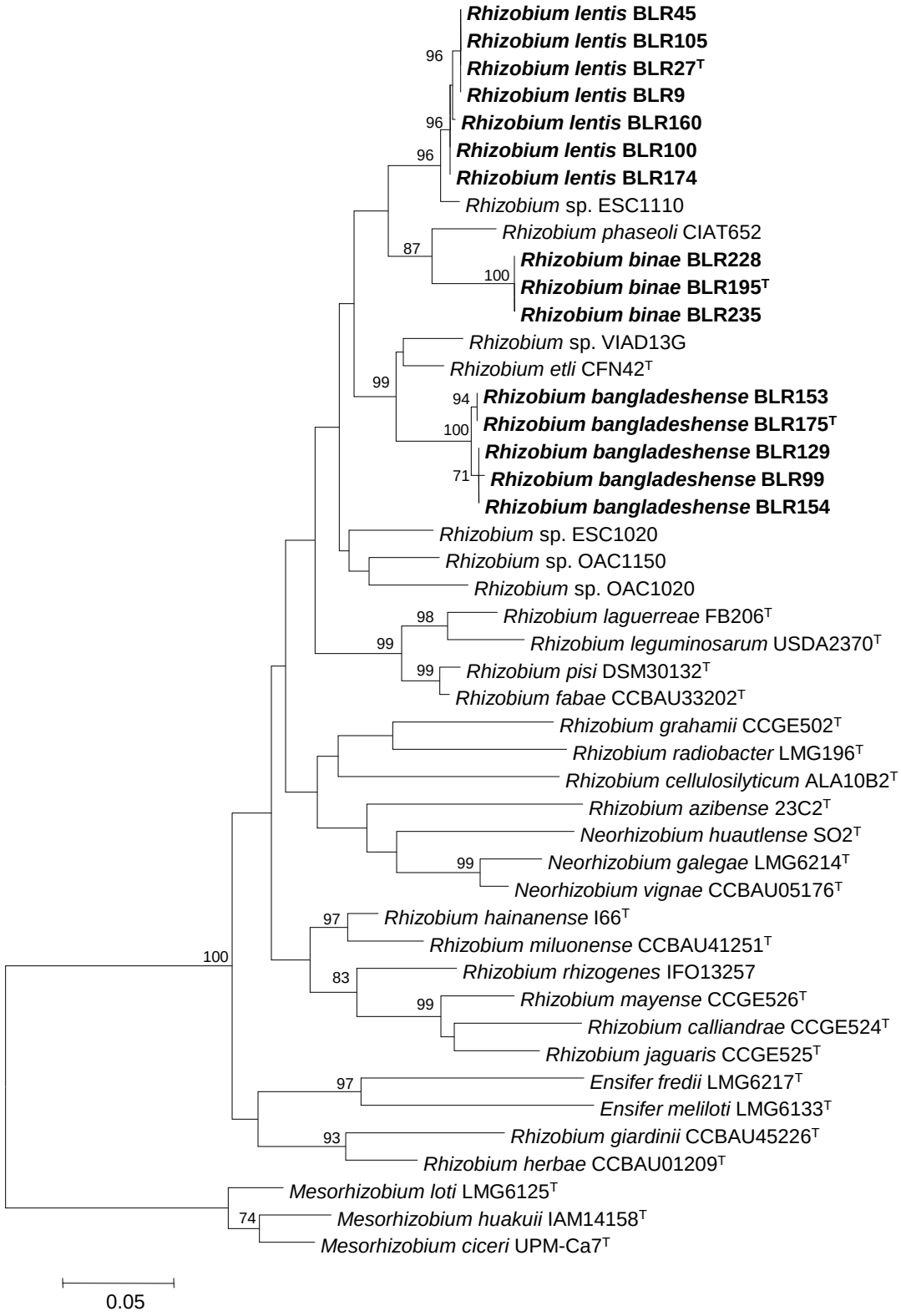
Rhizobium lentis BLR45
Rhizobium lentis BLR105
Rhizobium lentis BLR27T
Rhizobium lentis BLR9
Rhizobium lentis BLR160
Rhizobium lentis BLR100
Rhizobium lentis BLR174
Rhizobium sp. ESC1110
Rhizobium phaseoli CIAT652
Rhizobium binae BLR228
Rhizobium binae BLR195T
Rhizobium binae BLR235
Rhizobium sp. VIAD13G
Rhizobium etli CFN42T
Rhizobium bangladeshense BLR153
Rhizobium bangladeshense BLR175T
Rhizobium bangladeshense BLR129
Rhizobium bangladeshense BLR99
Rhizobium bangladeshense BLR154
Rhizobium sp. ESC1020
Rhizobium sp. OAC1150
Rhizobium sp. OAC1020
Rhizobium laguerreae FB206T
Rhizobium leguminosarum USDA2370T
Rhizobium pisi DSM30132T
Rhizobium fabae CCBAU33202T
Rhizobium grahamii CCGE502T
Rhizobium radiobacter LMG196T
Rhizobium cellulosilyticum ALA10B2T
Rhizobium azibense 23C2T
Neorhizobium huautlense SO2T
Neorhizobium galegae LMG6214T
Neorhizobium vignae CCBAU05176T
Rhizobium hainanense I66T
Rhizobium miluonense CCBAU41251T
Rhizobium rhizogenes IFO13257
Rhizobium mayense CCGE526T
Rhizobium calliandrae CCGE524T
Rhizobium jaguaris CCGE525T
Ensifer fredii LMG6217T
Ensifer meliloti LMG6133T
Rhizobium giardinii CCBAU45226T
Rhizobium herbae CCBAU01209T
Mesorhizobium loti LMG6125T
Mesorhizobium huakuii IAM14158T
Mesorhizobium ciceri UPM-Ca7T
96
96
96
87
100
99
94
100
71
98
99
99
97
100
83
99
97
93
74
0.05

Figure legends

Fig. 1. ML tree from 16S rRNA gene partial sequences. Bootstrap values ≥70 are indicated for each node (1000 replicates).

Fig. 2. ML tree from concatenated sequence of atpD-recA genes. Bootstrap values ≥70 are indicated for each node (1000 replicates).

International Journal of Systematic and Evolutionary Microbiology

Average nucleotide identity of genome sequences supports the description of *Rhizobium lentis* sp. nov., *Rhizobium bangladeshense* sp. nov. and *Rhizobium binae* sp. nov. from lentil (*Lens culinaris*) nodules

M. Harun-or Rashid[1,4*]. J. Peter W. Young[2], Isobel Everall[2], Pia Clercx[3], Anne Willems[3], Markus Santhosh Braun[4] and Michael Wink[4]

[1]Molecular Biology Laboratory, Soil Science Division, Bangladesh Institute of Nuclear agriculture (BINA), Bangladesh
[2]Department of Biology, University of York, UK
[3]Laboratory of Microbiology, Ghent University, Belgium
[4]Department of Biology, Institute of Pharmacy and Molecular Biotechnology, Heidelberg University, Germany

Corresponding author: Dr. M. Harun-or Rashid, Email: mhrashid08@gmail.com

Table S1. Genetic distances of type strains between proposed species and their close species

Species	Strains	Genetic distance (%)*					
		Rhizobium lentis		*Rhizobium bangladeshense*		*Rhizobium binae*	
		atpD	*recA*	*atpD*	*recA*	*atpD*	*recA*
Rhizobium lentis	BLR27^T	0.0	0.0	6.2	9.3	6.2	8.3
Rhizobium bangladeshense	BLR175^T	6.2	9.3	0.0	0.0	6.4	11.1
Rhizobium binae	BLR195^T	6.2	8.3	6.4	11.1	0.0	0.0
Rhizobium etli	CFN42^T	5.4	8.7	5.6	3.8	6.2	11.2
Rhizobium phaseoli	CIAT652	5.6	7.2	6.4	11.0	5.4	5.2
Rhizobium fabae	CCBAU33202^T	9.3	7.9	8.6	9.2	8.6	8.9
Rhizobium pisi	DSM30132^T	9.5	8.4	8.7	9.1	8.6	9.4
Rhizobium leguminosarum	USDA 2370^T	10.6	7.9	11.4	9.7	11.3	8.9

* “Genetic distance (%)*”: percentage of non-identical nucleotides”

Table S2. Genetic distances among the strains within species

Species	Genetic distances (%)*	
	Genes	
	atpD	*recA*
Rhizobium lentis (7 strains)	1.0	0.0
Rhizobium bangladeshense (5 strains)	0.4	0.5
Rhizobium binae (3 strains)	0.1	0.0

* “Genetic distance (%) ”: percentage of non-identical nucleotides”

Table S3. Additional differential substrate utilization patterns of the new species and closely related species. Data were obtained using Biolog GENIII systems according to the manufacturer's instructions. +, positive reaction, w, weak positive reaction, - negative reaction

	***R. bangladeshense* BLR 175^T**	***R. binae* BLR 195^T**	***R. lentis* BLR 27^T**	*R. etli* LMG 17827T	*R. fabae* LMG 23997^T	*R. laguerreae* LMG 27434^T	*R. leguminosarum* LMG 14904^T	*R. phaseoli* LMG 8819^T
Growth on:								
Dextrin	-	+	-	-	w	+	+	+
D-maltose, D-trehalose, D-cellobiose, gentiobiose, sucrose, D-raffinose, α-D-glucose	+	+	-	-	+	+	-	+
D-turanose	+	+	-	-	+	+	-	-
α-D-lactose, D-fructose	+	+	+	-	+	+	+	+
D-melibiose	w	+	-	w	+	+	+	+
β-methyl-D-glucoside, salicin	+	+	+	-	+	+	+	+
N-acetyl-D-mannosamine	w	-	-	-	-	+	+	-
N-acetyl-D-galactosamine	+	+	w	-	w	+	+	+
D-mannose	w	+	-	-	+	+	-	+
D-galactose	w	+	-	-	+	+	+	+
3-Methyl glucose	w	-	-	-	-	-	+	-
Inosine	w	-	-	-	-	-	-	+
D-sorbitol	+	+	+	-	+	+	-	+
D-mannitol	+	+	+	w	+	+	-	+
D-arabitol, glycerol	+	+	+	-	+	+	+	+
D-glucose-6-phosphate	+	+	w	w	-	-	+	+
D-fructose-6-phosphate	w	+	+	w	-	+	+	+
D-aspartic acid	-	w	-	-	-	+	-	+
Glycyl-L-proline	-	-	-	-	-	+	-	-
L-alanine	-	+	-	-	+	-	-	+
L-arginine	-	-	-	-	+	+	w	+

	***R. bangladeshense* BLR 175^T**	***R. binae* BLR 195^T**	***R. lentis* BLR 27^T**	*R. etli* LMG 17827T	*R. fabae* LMG 23997^T	*R. laguerreae* LMG 27434^T	*R. leguminosarum* LMG 14904^T	*R. phaseoli* LMG 8819^T
L-aspartic acid	w	+	+	-	+	+	-	+
L-glutamic acid	-	w	-	-	+	+	+	+
L-histidine	-	+	w	-	+	+	+	w
L-pyroglutamic acid	w	+	w	-	-	+	-	+
L-serine	-	w	-	-	-	+	-	+
Pectin	w	w	-	-	-	+	-	+
D-galacturonic acid	w	-	-	-	-	+	-	-
D-gluconic acid	+	w	+	-	-	+	+	+
D-glucuronic acid	w	-	w	w	-	+	-	w
Glucuronamide	w	-	w	+	w	+	+	w
Mucic acid	-	w	+	-	w	+	+	+
Quinic acid	+	+	w	-	+	+	w	+
D-Saccharic acid	+	+	-	-	-	+	-	+
p-Hydroxy-phenylacetic acid	-	-	-	-	-	+	-	-
Methyl pyruvate	-	+	-	-	+	+	-	-
D-lactic acid methyl ester	+	-	+	-	-	+	-	-
L-lactic acid	+	+	+	-	-	+	+	-
citric acid	-	+	-	-	+	+	+	-
α-keto-glutaric acid	+	-	w	w	-	+	+	-
D-malic acid, L-malic acid	-	+	+	-	+	+	+	+
Bromo-succinic acid	w	+	-	-	+	+	-	+
Tween 40	+	-	w	-	-	+	w	-
α-Hydroxy-butyric acid	w	w	w	-	w	+	w	-
β-Hydroxy-D,L-butyric acid	w	+	+	-	w	+	w	w
Acetoacetic acid	w	w	-	-	+	+	-	-
Propionic acid	-	-	-	-	+	+	+	-
Acetic acid	w	+	+	-	+	+	+	+

	***R. bangladeshense* BLR 175^T**	***R. binae* BLR 195^T**	***R. lentis* BLR 27^T**	*R. etli* LMG 17827T	*R. fabae* LMG 23997^T	*R. laguerreae* LMG 27434^T	*R. leguminosarum* LMG 14904^T	*R. phaseoli* LMG 8819^T
Formic acid	-	-	+	-	+	+	w	-
Growth at pH 6	-	+	-	-	w	+	-	-
Growth in the presence of:								
1% NaCl	-	-	-	-	w	+	-	-
1% Sodium lactate	-	-	-	-	-	+	-	-
Troleandomycin	-	-	-	-	-	+	+	-
Rifamicin SV	w	+	+	-	+	+	+	+
Lincomycin	+	+	+	+	+	+	+	-
Tetrazolium violet	-	w	+	+	w	-	+	+
Tetrazolium blue	-	w	+	+	-	-	+	+
Nalidixic acid	-	+	+	+	+	+	+	w
Lithium chloride	-	-	-	-	-	+	-	-
Pottasium tellurite	+	+	-	+	+	+	+	+
Sodium butyrate	-	-	-	+	-	-	-	-

In Biolog III systems, all strains were able to grow on D-fucose, L-fucose and L-rhamnose and none of the strains was able to use D-serine, D-galactonic acid lactone or α-keto-butyric acid. None of the strains could grow at pH 5 or in the presence of 4% or 8% NaCl, fusidic acid, D-serine, minocycline, guanidine HCl, niaproof 4, vancomycin, aztreonam or sodium bromate in Biolog III systems.

Table S4. Fatty acid pattern of type strain of proposed three new species.

Fatty acid	*Rhizobium lentis* (BLR27^{T})	*Rhizobium bangladeshense* (BLR175^{T})	*Rhizobium binae* (BLR195^{T})
15:0 iso 2 OH	0.55	absent	0.79
16:0	2.03	2.67	2.12
16:0 3 OH	0.91	2.16	0.80
18:0	3.23	6.88	4.63
18:1ω9c	absent	absent	2.36
18:0 3 OH	absent	absent	0.97
18:1ω7c 11methyl	absent	absent	0.63
19:0 cyclo ω8c	11.37	6.66	21.76
*Summed featured 2	5.80	7.00	5.62
†Summed featured 3	0.43	absent	absent
‡Summed featured 8	75.35	74.62	60.32

Summed featured indicate two or more fatty acids that could not be separated by MIDI system.
*Summed featured 2 consist one or more of 12:0 (aldehyde), unknown ECL 15.489, 14:0 3 OH / 16:1 iso I.
†Summed featured 3 consists 16:1ω7c / 16:1ω6c.
‡Summed featured 8 consists 18:1ω7c / 18:1ω6c
NB: Fatty acids present in ‹0.5% are not mentioned in this table.

Table S5. Phenotypic characteristics of proposed three new species and their close relatives

Species	NaCl tolerance (%)			Temperature tolerance (^{0}C)	pH tolerance		Antibiotic sensitivity (µg mL^{-1})						
							Ampicillin					Kanamycin	
	0.5	1.0	2.0	37	8.2	10	50	75	100	125	150	10	20
R. lentis (21 strains)	—	—	—	+	+	v	+	+	+	v	v	+	+
R. bangladeshense (6 strains)	+	—	—	+	+	+	+	v	v	v	v	+	+
R. binae (3 strains)	+	v	—	+	+	+	+	—	—	—	—	+	v
*R. etli**	—	—	ND	+	—	—	—	—	—	—	—	—	—
*R. phaseoli**	+	—	—	v	—	—	+	+	+	+	+	ND	ND
*R. pisi**	+	+	—	+	+	—	—	—	—	—	—	ND	ND
*R. fabae**	+	+	+	+	+	—	+	+	+	+	+	+	+
*R. leguminosarum**	+	—	—	—	+	ND	v	v	v	v	v	ND	ND
*R. laguerreae**	+	—	—	+	ND	ND	v	v	v	v	v	ND	ND

Abbreviations: + = growth positive for >80% of strains, — = growth negative for >80% of strains, v = strains varied in response and w= weak response. ND = not determined / data not available, *R.* = *Rhizobium*.

* Data taken from Rashid *et al*. (2012) and published species descriptions (Ramírez-Bahena *et al*., (2008); Saïdi *et al*., (2014); Segovia *et al*., (1993).

N.B.: All strains grew at 32^{0}C and showed acidic reaction on BTB medium, but none grew at 40^{0}C temperature, in tetracycline (5 µg mL^{-1}) or on LB medium.