

SUPPLEMENTARY MATERIAL

***Argyrolobium* legumes from an African centre of endemism associates with novel *Bradyrhizobium* species harbouring unique sets of symbiosis genes**

Mabodiba M. Maake¹, Chrizelle W. Beukes^{2,1}, Magriet A. van der Nest^{1,3}, Juanita R. Avontuur⁴, Esther K. Muema⁵, Tomasz Stępkowski⁶, Stephanus N. Venter¹, Emma T. Steenkamp¹

¹ Department of Biochemistry, Genetics, and Microbiology, Forestry and Agricultural Biotechnology Institute, University of Pretoria, South Africa

² Department of Ecological Sciences, The James Hutton Institute, Invergowrie, Dundee, DD2 5DA, UK

³ Hans Merensky Chair in Avocado Research, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Pretoria 0002, South Africa

⁴ Marine Microbiology Group (MMG), Department of Animal and Microbial Biodiversity, Mediterranean Institute for Advanced Studies (IMEDEA, CSIC-UIB), Esporles, Spain.

⁵ Department of Soil Science, Faculty of AgriSciences, Stellenbosch University, South Africa

⁶ Department of Biochemistry and Microbiology, Institute of Biology, Warsaw University of Life Sciences (SGGW), Poland

corresponding authors:emma.steenkamp.co.za and mabodibamaake@gmail.com

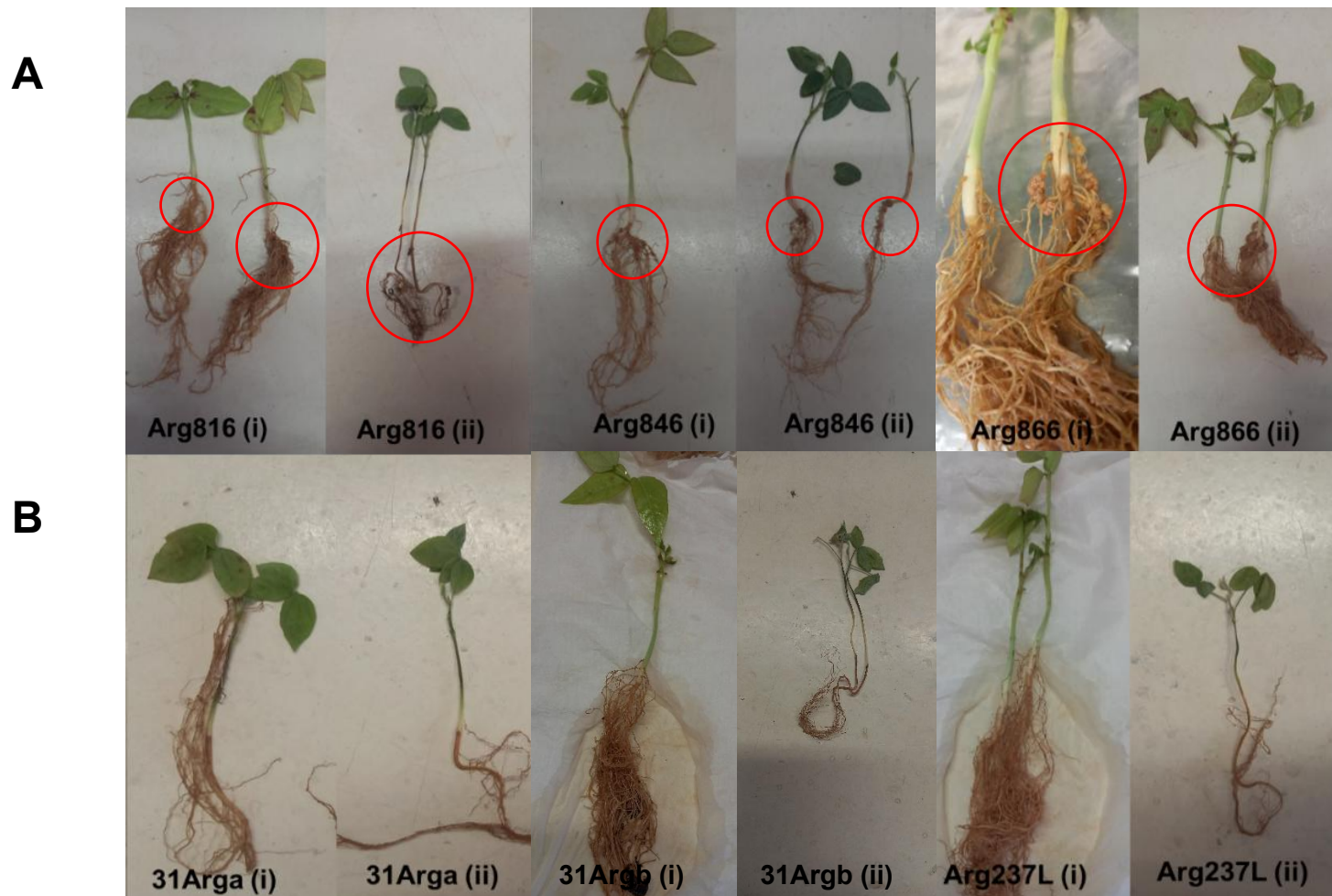


Figure S1. Glasshouse nodulation experiments conducted with the hosts (i) cowpea and (ii) siratro. Two seeds were planted per pot and tests were done in duplicate.

B. spitzkopense isolates (A) and *B. mpumalangense* isolates (B). Nodules were only observed for strains Arg816^{Ts}, Arg846, and Arg866 (circled in red), with hosts inoculated with isolate Arg866 showing a higher number of nodules followed by Arg846. Nodule morphology was spherical with pink interiors.

Figure S2. A maximum likelihood phylogeny of the *dnaK* dataset for *Bradyrhizobium*. Phylogenies were constructed using 1000 bootstrap iterations in RAxML v. 8.2.1 (Stamatakis, 2014). Only bootstrap support of $\geq 70\%$ are indicated. *Rhodopseudomonas* (strains: HaA2 and CGA009) was used for outgroup purposes. The scale bar represents the number of nucleotide changes per site. Strains from this study are in bold and coloured in pink (*B. spitzkopense* strains), yellow (*B. mpumalangense* strains), red and blue (conspecific strains). For all the species the type strain number is listed after the name. Sequence accession numbers obtained from the GenBank (National Centre for Biotechnology Information (NCBI); <http://www.ncbi.nlm.nih.gov/>; Benson et al., 2004) and isolation sources are shown in Table S2 and S3.

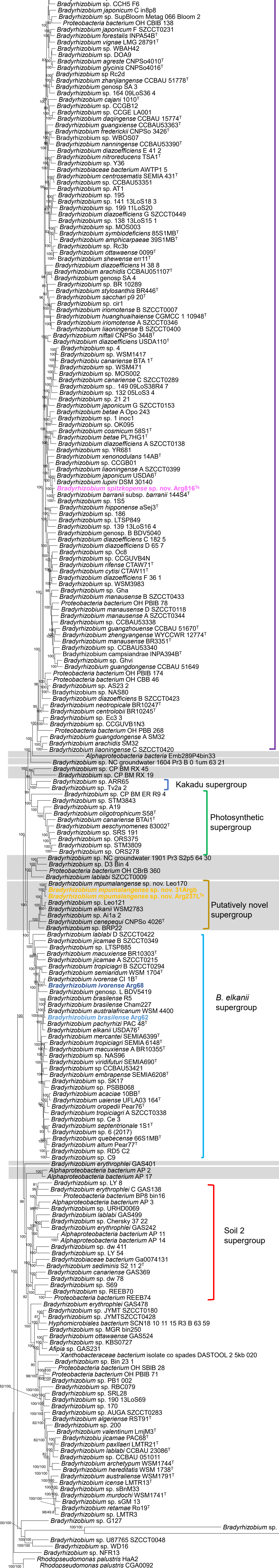
Figure S3. A maximum likelihood phylogeny of the *glnII* dataset for *Bradyrhizobium*. Phylogenies were constructed using 1000 bootstrap iterations in RAxML v. 8.2.1 (Stamatakis, 2014). Only Bootstrap support of $\geq 70\%$ are indicated. *Rhodopseudomonas* (strains: HaA2, CGA009, and BisB5) was used for outgroup purposes. The scale bar represents the number of nucleotide changes per site. Strains from this study are in bold and colored in pink (*B. spitzkopense* strains), yellow (*B. mpumalangense* strains), red and blue (conspecific strains). For all the species the type strain number is listed after the name. Sequence accession numbers obtained from the GenBank (National Centre for Biotechnology Information (NCBI); <http://www.ncbi.nlm.nih.gov/>; Benson et al., 2004) are shown in Suppl. Table S3.

Figure S4. A maximum likelihood phylogeny of the *gyrB* dataset for *Bradyrhizobium*. Phylogenies were constructed using 1000 bootstrap iterations in RAxML v. 8.2.1 (Stamatakis, 2014). Only bootstrap support of $\geq 70\%$ are indicated. *Rhodopseudomonas* (strains: HaA2, CGA009, BisB5) was used for outgroup purposes. The scale bar represents the number of nucleotide changes per site. Strains from this study are in bold and coloured pink (*B. spitzkopense* strains), yellow (*B. mpumalangense* strains), red and blue (conspecific strains). Sequence accession numbers obtained from the GenBank (National Center for Biotechnology Information (NCBI); <http://www.ncbi.nlm.nih.gov/>; Benson et al., 2004) are shown in Suppl. Table S3.

Figure S5. A maximum likelihood phylogeny of the *recA* dataset for *Bradyrhizobium*. Phylogenies were constructed using 1000 Ultrafast bootstrap iterations RAxML v. 8.2.1 (Stamatakis, 2014). Only bootstrap support of $\geq 70\%$ are indicated. *Rhodopseudomonas* (strains: HaA2, CGA009, and BisB5) was used for outgroup purposes. The scale bar represents the number of nucleotide changes per site. Strains from this study are in bold and coloured in pink (*B. spitzkopense* strains), yellow (*B. mpumalangense* strains), red and blue (conspecific strains). For all the species the type strain number is listed after the name. Sequence accession numbers obtained from the GenBank (National Centre for Biotechnology Information (NCBI); <http://www.ncbi.nlm.nih.gov/>; Benson et al., 2004) are shown in Suppl. Table S3.

Figure S6. A maximum likelihood phylogeny of the *rpoB* dataset for *Bradyrhizobium*. Phylogeny was constructed using 1000 bootstrap iterations in RAxML v. 8.2.1 (Stamatakis, 2014). Only bootstrap support of $\geq 70\%$ is indicated. *Rhodopseudomonas* (strains: HaA2, CGA009) was used for outgroup purposes. The scale bar represents the number of nucleotide changes per site. Strains from this study are in bold and coloured in pink (*B. spitzkopense* strains), yellow (*B. mpumalangense* strains), red and blue (conspecific strains). For all the species the type strain number is listed after the name. Sequence accession numbers obtained from the GenBank (National Centre for Biotechnology Information (NCBI); <http://www.ncbi.nlm.nih.gov/>; Benson et al., 2004) are shown in Suppl. Table S3.

UBCG
(92-core genes)



B. japonicum
supergroup

Kakadu supergroup

Photosynthetic
supergroup

Putatively novel
supergroup

B. elkanii
supergroup

Soil 2
supergroup

B. jicamae
supergroup

Bradyrhizobium sp. LY 7
Bradyrhizobium lupini HPC L

Figure S7. A 92 gene, 250 taxa phylogenomic IQ-TREE following the Up-to-date Bacterial Core Gene (UBCG). Dataset comprised representative *Bradyrhizobium* genomes from the Genome Taxonomy Database (GTDB), together with strains previously described as *B. brasilense* (Arg62, R5, cham227, and UFLA 03321^T) and described *Bradyrhizobium* species with available genomes. Strains from this study are in bold and colored in pink (*B. spitzkopense* strains), yellow (*B. mpumalangense* strains), red and blue (conspecific strains). UFBoot and SH-aLRT support are indicated at the nodes. *Rhodopseudomonas* strains HaA2 and CGA0092 were used as outgroups. The scale bar represents the number of nucleotide changes per site.

Clade V Photosynthetic

Clade XII Nepal

Clade VIII tropical & subtropical Americas

Clade VII tropical and subtropical Americas + Japan

Cosmopolitan Clade III

Clade X Australia

Clade XI Spain

Clade IX Zimbabwe
Clade XVI Central America

Clade V Tropical and Subtropical Americas

Clade XIV South Africa

Clade XIII South Africa

Clade XV South Africa

Clade I Australia

Clade X Australia

Clade IV Australia + Spain +
Morocco +China

Clade II Europe &
Mediterranean

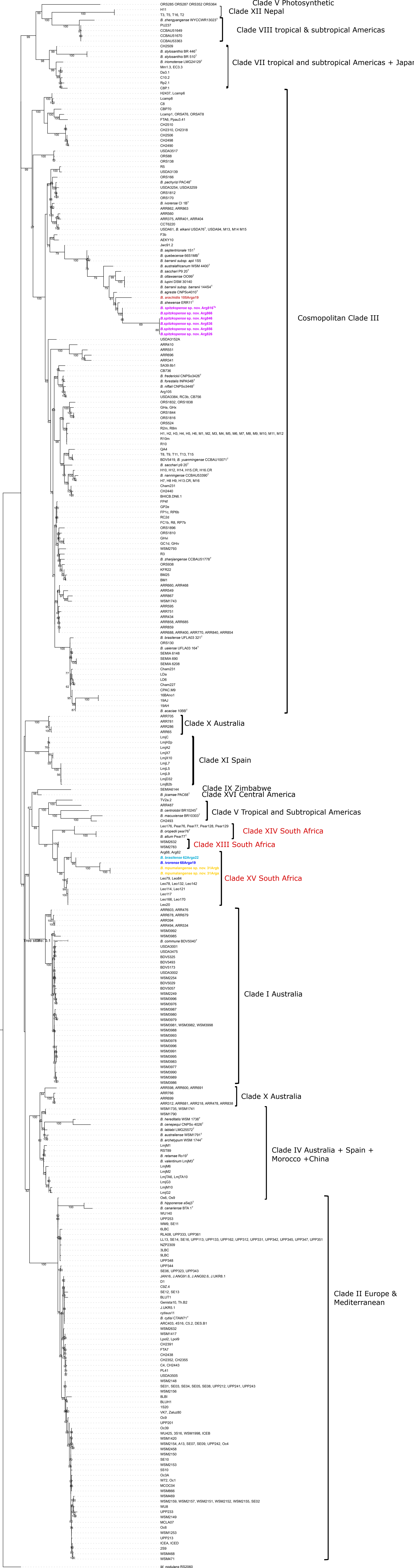


Figure S8. *nodA* maximum likelihood phylogenetic tree based on 244 haplotypes and 44 additional *Bradyrhizobium nodA* amino acid sequences aligned sequences of 642 bp. *Methylobacterium nodulans* ORS 2060 was used for outgroup purposes strains from this study are in bold and highlighted. GenBank accession numbers are listed in Suppl. Table S2 of the Beukes et al. (2016) study and Suppl. Table S5 of this study. Bootstrap values 60 are indicated at the nodes. The scale bar represents the number of nucleotide changes per site.

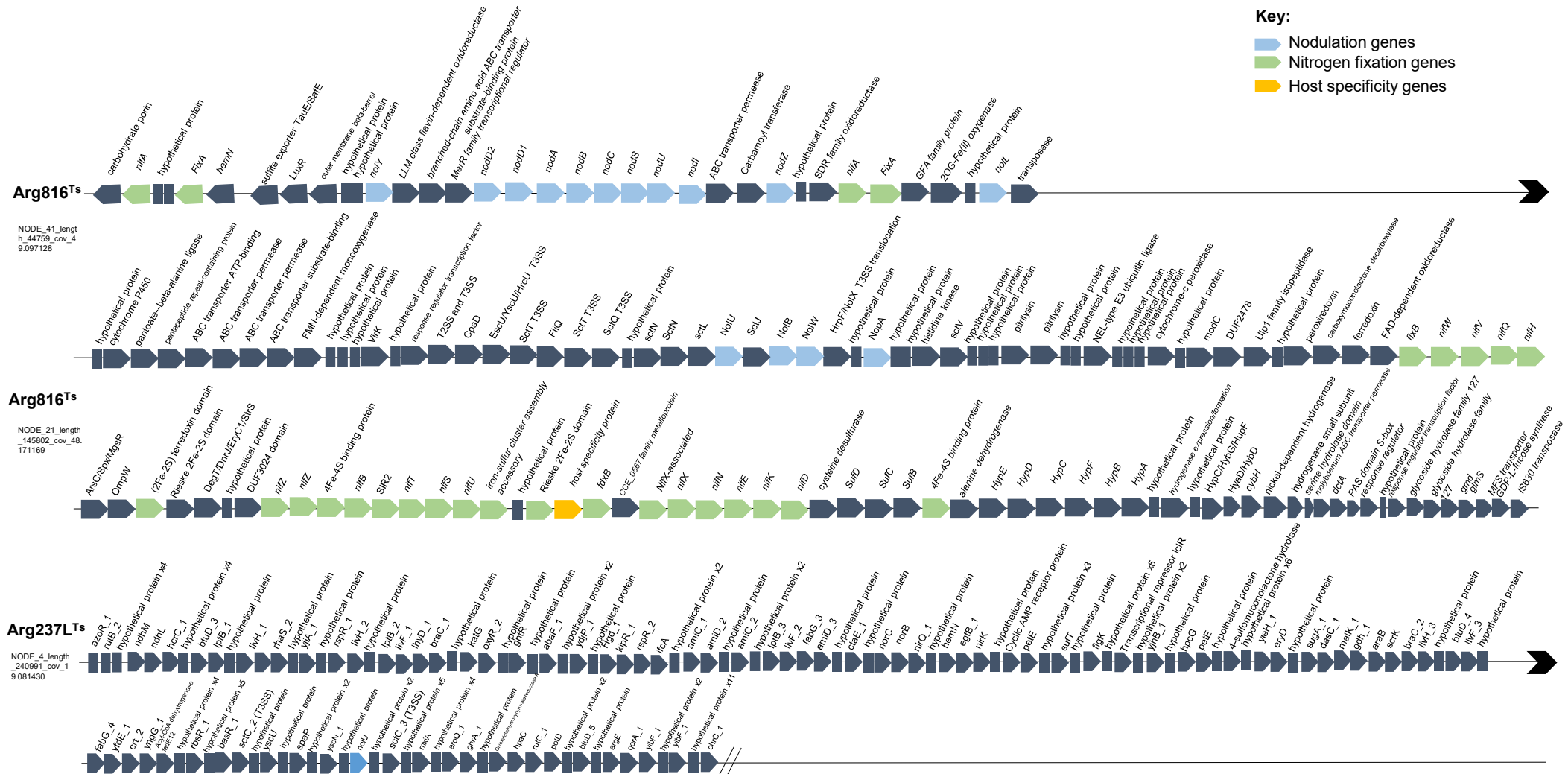
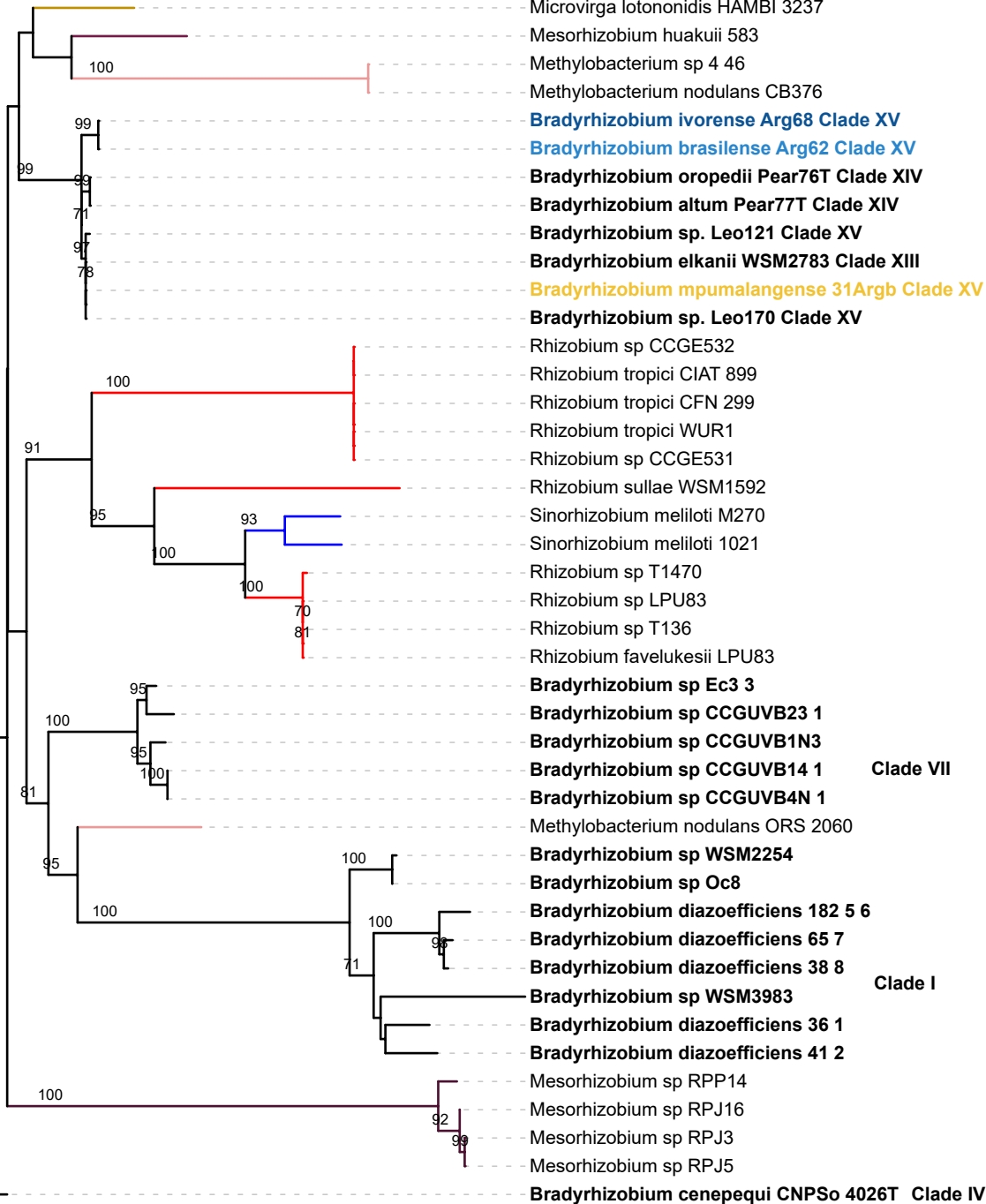


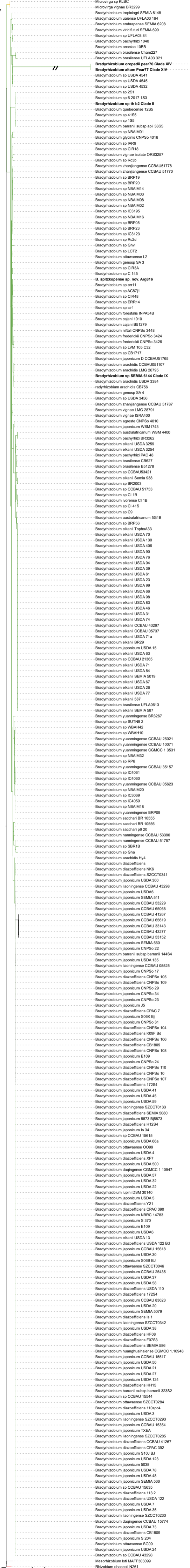
Figure S10. Schematic diagram showing genetic map of contigs containing the *nod* and *nif* clusters in the type strains Arg816^T and Arg237L^T. Genes were screened from a Prokka annotation and NCBI annotation protein sequence search. In Arg237L^T, three *nodD* and two *nodM* homologs were found on distantly from each other upstream of the contig shown here (i.e., *nodD2_1* on contig 5 at position 851054 bp, *nodM_1* on contig 6 at 1184406 bp, *nodD_2* on contig 35 at position 4814657 bp, *nodD_3* at position 6167210 bp, *nodM_2* on contig 64 at position 6872077 bp. Additionally in Arg237L^T, *nolB*, *nolW*, *nolU*, *nfeD* genes (accessions: WP_282526875.1, WP_282526876.1, WP_282526873.1, WP_282527884.1) were additionally found based on NCBI reseq annotation.

nodH



Tree scale: 0.1

Figure S11. *nodH* based Maximum likelihood phylogeny constructed from 43 taxa, using *B. cenepequi* CNPSo 2046 as the outgroup. Nodes display ultrafast bootstraps followed by SH-aLRLT. The scale bar represents the number of nucleotide changes per site.



Tree scale: 1

Figure S12. *nodZ* based Maximum likelihood phylogeny constructed from 275 taxa with *Mesorhizobium loti* MAFF303099 and *Rhizobium phaseoli* N261 used as outgroups.

noeE

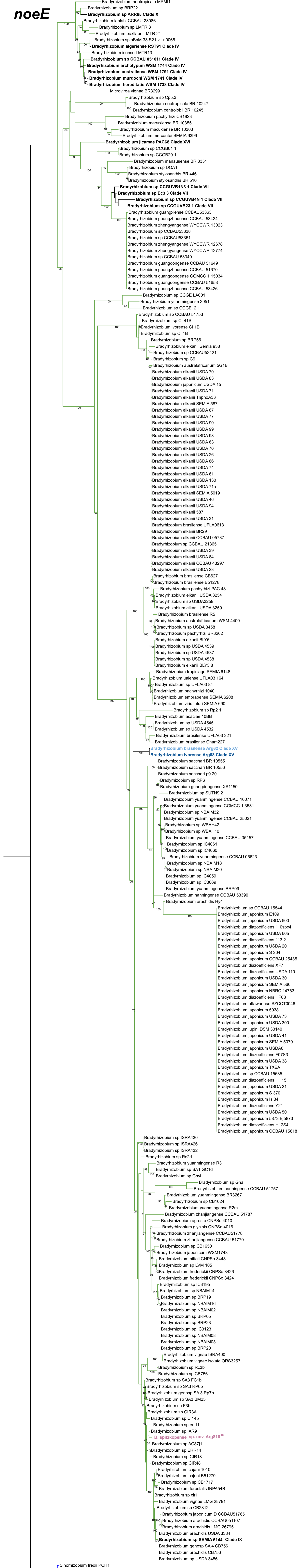


Figure S13. *noeE* Maximum likelihood phylogeny constructed from 247 taxa with *Sinorhizobium fredii* strains NGR234 and PCH1 used as outgroups

noel

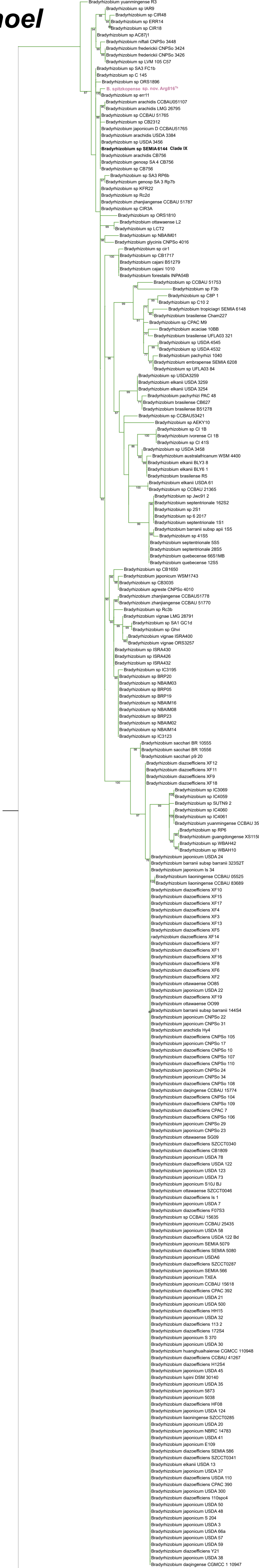


Figure S14. *noel* based Maximum likelihood phylogeny constructed from 237 taxa with *Rhizobium etli* CIAT 652 and *Sinorhizobium fredii* PCH1 used as outgroups