

Draft genome sequences of four potential new species of the genus *Bradyrhizobium* isolated from root nodules of native legumes in Costa Rican forests

Rachelle Fernández-Vargas^{1,2}, Sara Jiménez-Alpízar^{1,3}, Valeria Leandro-Arce^{1,3,4}, Bradd Mendoza-Guido^{5,*} and Keilor Rojas-Jimenez^{1,*}

Abstract

Here, we report the draft genome sequences of four *Bradyrhizobium* spp. isolates obtained from root nodules of the native legumes *Pentaclethra macroloba*, *Chamaecrista nictitans*, *Erythrina fusca* and *Zygia engelsingii* in tropical forests of Costa Rica. Genomes ranged from 8.6 to 9.8 Mb with GC contents between 62.8% and 63.8%. Phylogenomic analysis, along with average nucleotide identity (ANI) and digital DNA–DNA hybridization (dDDH) comparisons, confirmed that these isolates represent potential new species. ANI values ranged from 88.3% to 90.3%, and dDDH values from 28.8% to 41.8%, compared with their closest *Bradyrhizobium* species. Functional annotation revealed some genes related to nitrogen fixation (*nifA*, *nifB*, *nifH*) and nodulation capacity (*nodB*, *nodC*, *nodJ*). These results provide insights into the diversity and symbiotic capabilities of *Bradyrhizobium* in tropical ecosystems.

DATA SUMMARY

The raw read sequences of the four *Bradyrhizobium* isolates have been deposited in GenBank, under the BioProject accession no. PRJNA1084876 and Sequence Read Archive (SRA) accession no. SRP493729. The sample accession details are as follows: *Bradyrhizobium* sp. LEGMi-117d (BioSample: SAMN44800723, SRA: SRR31439459, 16S rRNA: PP439200), *Bradyrhizobium* sp. LEGMi-161a (BioSample: SAMN44800724, SRA: SRR31439458, 16S rRNA: PP439223), *Bradyrhizobium* sp. LEGMi-163b (BioSample: SAMN44800725, SRA: SRR31439457, 16S rRNA: PP439225) and *Bradyrhizobium* sp. LEGMi-207a (BioSample: SAMN44800726, SRA: SRR31439456, 16S rRNA: PP439252).

ANNOUNCEMENT

Legumes have established symbiotic relationships with nitrogen-fixing bacteria, enabling them to thrive in nutrient-poor soils and contributing significantly to ecosystem sustainability [1, 2]. Among these bacteria, *Bradyrhizobium* (*Nitrobacteraceae*, *Alphaproteobacteria*) is a key genus involved in symbiotic nitrogen fixation, exhibiting notable diversity, particularly in tropical environments [3, 4]. However, knowledge about its genomic diversity, particularly in tropical regions and less-studied legumes, remains limited. To help address this gap, we analysed *Bradyrhizobium* isolates obtained from nodules of native leguminous plants in Costa Rica, a region with rich but underexplored microbial diversity.

In this work, we characterized four *Bradyrhizobium* isolates obtained from *Fabaceae* plants in Costa Rica. The four native *Fabaceae* species from tropical Costa Rican forests are as follows: *Pentaclethra macroloba*, *Chamaecrista nictitans*, *Erythrina fusca* and

Access Microbiology is an open research platform. Pre-prints, peer review reports, and editorial decisions can be found with the online version of this article. Received 27 January 2025; Accepted 02 May 2025; Published 16 May 2025

Author affiliations: ¹Escuela de Biología, Universidad Costa Rica, San José, Costa Rica; ²Programa de Posgrado en Microbiología, Parasitología, Química Clínica e Inmunología (PPMPGCI), Universidad Costa Rica, San José, Costa Rica; ³Programa de Posgrado en Biología (PPB), Universidad de Costa Rica, San José, Costa Rica; ⁴Centro Nacional de Innovaciones Biotecnológicas (CENIBiot), CeNAT-CONARE, San José, Costa Rica; ⁵Instituto de Investigaciones en Salud, Universidad de Costa Rica, San José, Costa Rica.

***Correspondence:** Bradd Mendoza-Guido, bradd.mendoza@ucr.ac.cr; Keilor Rojas-Jimenez, Keilor.rojas@ucr.ac.cr

Keywords: bacteria; *Fabaceae*; nitrogen fixation; nodulation; *Rhizobium*; symbiosis.

Abbreviations: ANI, average nucleotide identity; dDDH, digital DNA–DNA hybridization; iTOL, interactive tree of life; PY, peptone yeast.

Sequence data were deposited at the NCBI Sequence Read Archive under accession number PRJNA1084876.

000990.v3 © 2025 The Authors



Table 1. Genome characteristics of the four *Bradyrhizobium* spp. isolates analysed in this work

Characteristic(s)	Isolates			
	<i>Bradyrhizobium</i> sp. LEGMi-117d	<i>Bradyrhizobium</i> sp. LEGMi-161a	<i>Bradyrhizobium</i> sp. LEGMi-163b	<i>Bradyrhizobium</i> sp. LEGMi-207a
Host species	<i>Pentaclethra macroloba</i>	<i>Chamaecrista nictitans</i>	<i>Erythrina fusca</i>	<i>Zygia engelsingii</i>
Completeness (%)	100	100	100	100
Coverage (x)	127	161	143	158.5
No. Contigs	93	55	60	75
N50 (pb)	359,180	492,614	600,185	417,354
Genome length (bp)	9,865,395	8,621,444	9,007,925	9,226,589
CDS	9216	8079	8341	8466
GC content (%)	62.79	63.63	63.81	63.13
<i>nod</i> genes	<i>nodS</i>	<i>nodA, noB, nodS, nodIJ</i>	<i>nodA, nodB, nodIJ, nodS</i>	<i>nodA, nodB, nodC, nodD1, nodIJ, nodS, nodZ</i>
<i>nif</i> genes	<i>nifA, nifB, nifD_1, nifD_2, nifK_1, nifK_2, nifS_1, nifS_2, nifW2, calB</i>	<i>nifA, nifB, nifD, nifH_1, nifH_2, nifK_1, nifK_2, nifS, nifW</i>	<i>nifA, nifB, nifD, nifH, nifK_1, nifK_2, nifS_1, nifS_2, nifW, calB</i>	<i>nifA, nifB, nifD, nifH_1, nifH_2, nifK_1, nifK_2, nifS_1, nifS_2, nifW2, calB</i>

Table 2. ANI values between *Bradyrhizobium* strains under study and their closest validly published species

	<i>Bradyrhizobium</i> <i>hipponense</i> aSej3 ^T GCF_008123965.1	<i>Bradyrhizobium</i> <i>yuanmingense</i> CCBAU 10071 ^T GCF 900094575 1	<i>Bradyrhizobium</i> <i>centrolobii</i> BR 10245 T ^T GCF_001641635.1	<i>Bradyrhizobium</i> <i>barranii</i> subsp. <i>barranii</i> 144S4 ^T GCA_017565645.3
<i>Bradyrhizobium</i> sp. LEGMi-161a	86.05	92.11	85.85	87.15
<i>Bradyrhizobium</i> sp. LEGMi-207a	83.82	83.51	84.32	83.85
<i>Bradyrhizobium</i> sp. LEGMi-163b	86.86	87.18	86.44	90.13
<i>Bradyrhizobium</i> sp. LEGMi-117d	88.26	86.24	87.23	87.78

Zygia engelsingii. The isolation and DNA extraction processes were performed following the protocol described by Rojas-Jimenez et al. [5]. Briefly, root nodules were collected from young trees (30–100 cm in height, 1–3 years old) from three nurseries across different regions of Costa Rica. Such trees are regularly used in reforestation programmes at the local level, using soil preparations from these regions without any pre-inoculation. The nodules were surface-sterilized using sequential treatments with 70% ethanol and 2% sodium hypochlorite for 5 min, followed by rinsing with sterile distilled water. After maceration in sterile PBS (pH 7.4), bacterial suspensions were serially diluted (10^{-1} and 10^{-2}) in peptone-yeast (PY) extract broth (5 g l⁻¹ peptone, 1 g l⁻¹ anhydrous dextrose, 0.5 g l⁻¹ dipotassium phosphate, 0.2 g l⁻¹ magnesium sulphate, 0.1 g l⁻¹ sodium chloride), plated on PY-agar with cycloheximide to inhibit fungal growth and incubated at 28 °C for up to 3 weeks. Bacterial colonies were purified through streaking and cryopreserved in glycerol stocks at –80 °C.

DNA extraction from the four bacterial isolates was performed using the DNEasy PowerSoil Kit (Qiagen, USA), following manufacturer's instructions. Sequencing was conducted with a NovaSeq PE150 platform (Illumina) at Novogene Co. (CA, USA)

Table 3. dDDH (d4) values (in %) between *Bradyrhizobium* strains under study and their closest validly published species

	<i>B. hippoense</i> aSej3 ^T GCF_008123965.1	<i>B. yuanmingense</i> CCBAU 10071 ^T GCF 900094575 1	<i>B. centrolobii</i> BR 10245 T ^T GCF_001641635.1	<i>B. barranii</i> subsp. <i>barranii</i> 144S4 ^T GCA_017565645.3
<i>Bradyrhizobium</i> sp. LEGMi-161a	31.40	48.40	31.30	35.30
<i>Bradyrhizobium</i> sp. LEGMi-207a	31.30	27.50	28.80	28.30
<i>Bradyrhizobium</i> sp. LEGMi-163b	33.00	33.80	39.50	41.30
<i>Bradyrhizobium</i> sp. LEGMi-117d	36.50	32.00	34.00	35.50

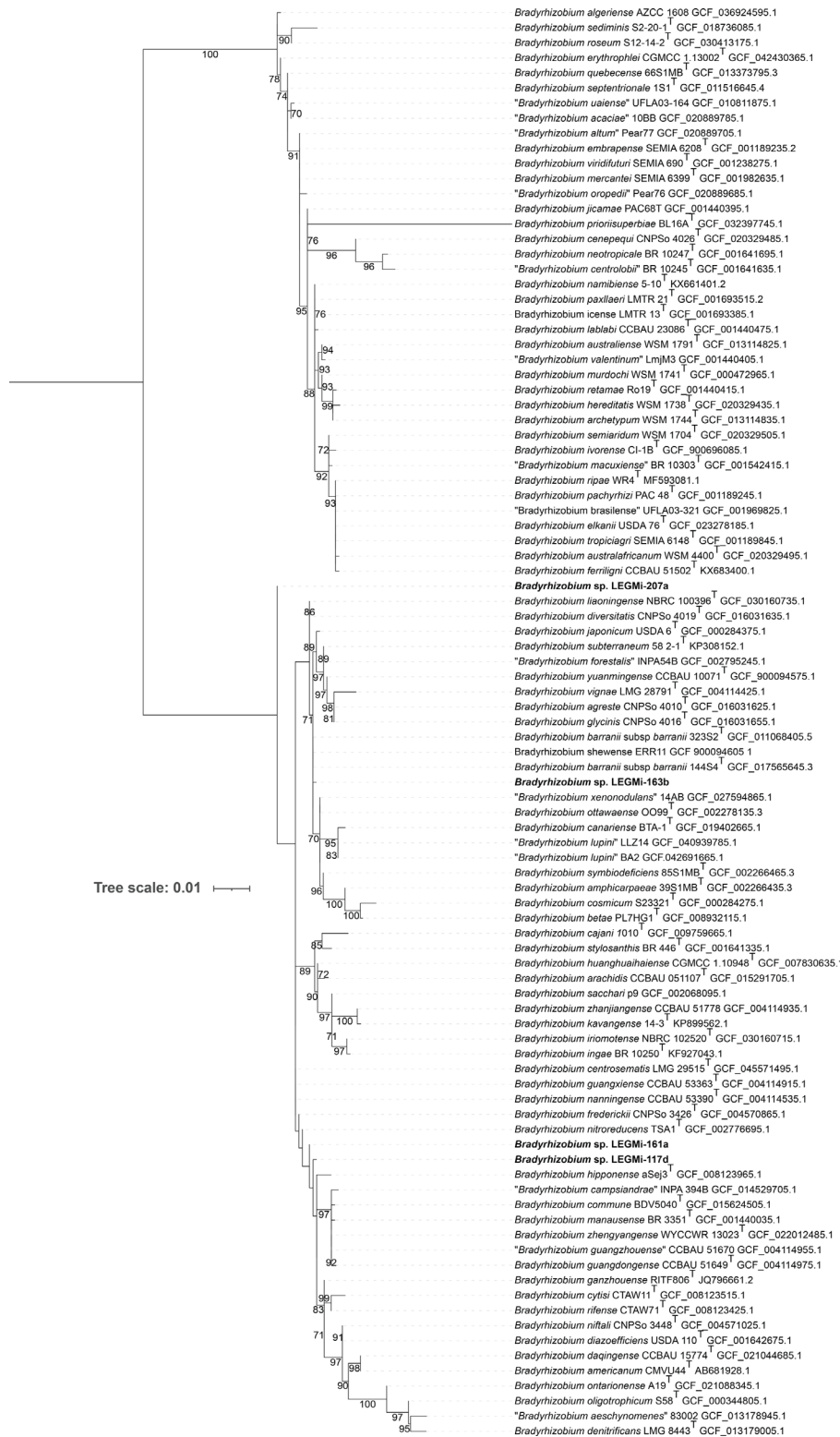


Fig. 1. Maximum-likelihood 16S rRNA gene phylogenetic tree of *Bradyrhizobium* strains including four putative novel species. The phylogeny was generated with IQ-TREE v2 under the GTR+G+I nucleotide substitution model with 1,000 bootstrap replicates. The tree was visualized using iTOL v6. Reference genomes included in the tree were manually selected from the GenBank database. Accession numbers are provided next to each strain; for strains without a genome accession, the 16S rRNA gene accession number is shown instead. Bootstrap values $\geq 70\%$ are shown at the nodes. The four putative new *Bradyrhizobium* species are highlighted in bold.

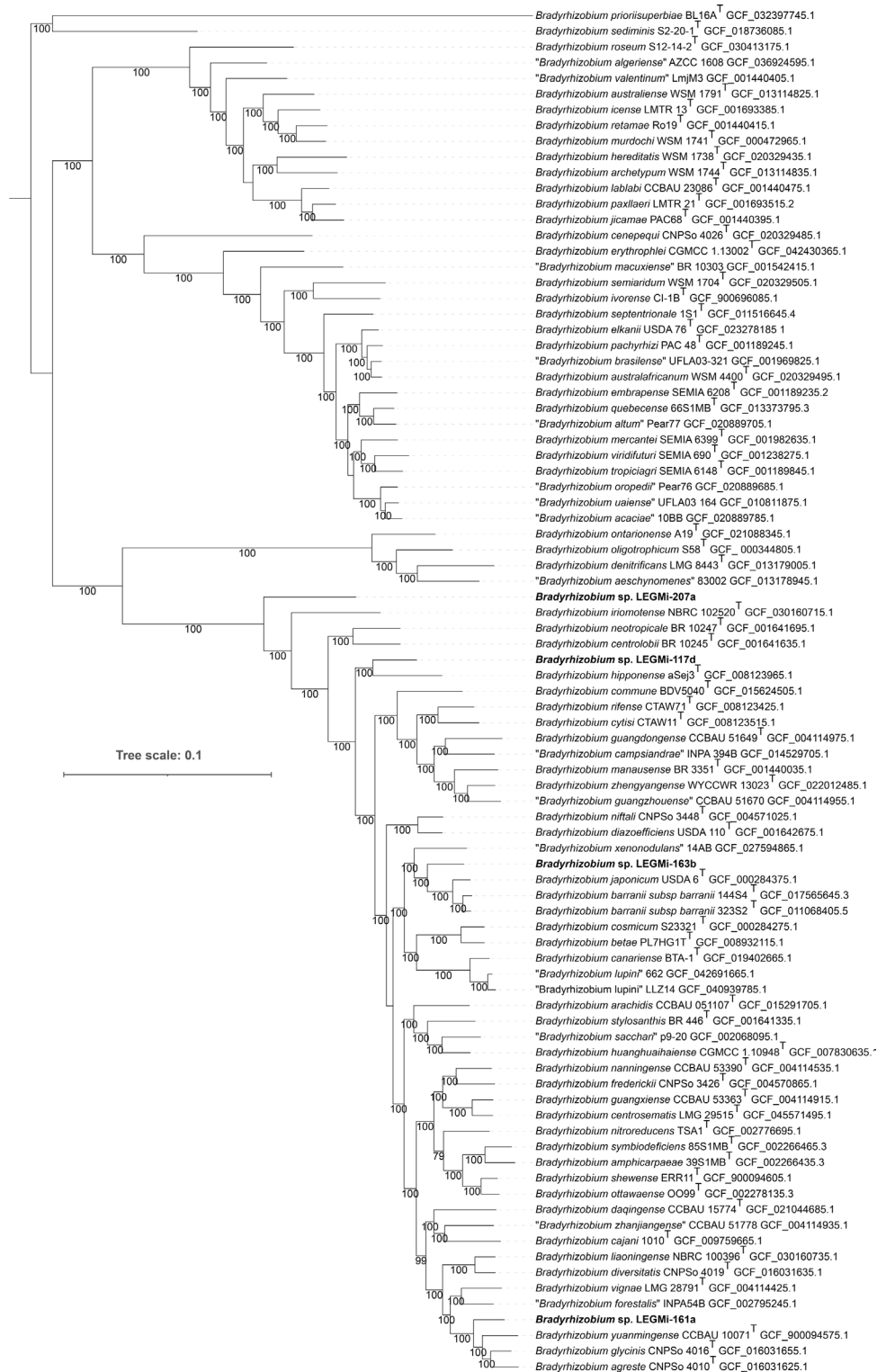


Fig. 2. Phylogenomic tree of *Bradyrhizobium* spp. highlighting four potential new species obtained in this study. The phylogeny was constructed using FastTree [18] with the JTT+CAT model and 1,000 bootstrap replicates for robustness. A total of 745 single-copy orthologous gene sequences were used to build the concatenated alignment for tree inference. Visualization was performed using iTOL v6 [19], and the tree was rooted at midpoint. Reference genomes included in the tree were manually selected from the GenBank database, prioritizing type strains closely related to the query genomes. Accession numbers for each genome are indicated in labels, and bootstrap values superior to 70% are displayed at the nodes. The four putative new *Bradyrhizobium* species are highlighted in bold.

for all strains. Additionally, the genome of isolate 163b was sequenced using the PacBio RS II platform (Pacific Biosciences). Quality assessment and filtering of raw sequencing data were performed with fastp v0.20.1 (quality score >25) [6]. High-quality Illumina reads were assembled using SPAdes v3.15.4 [7]. For isolate 163b, a hybrid assembly was also performed using the same tool. Both assemblies were generated with the following k-mer values: 33, 55, 77, 99 and 127. Contigs shorter than 1,000 bp were filtered out using Seqtk v1.2 [8], assembly statistics were generated with QUAST v5.2.1 [9], completeness and redundancy of the assembled genomes were evaluated with CheckM2 v1.0.1 [10], and functional annotation of the genomes was performed using eggNOG-mapper v2 [11] (Table 1).

We manually compiled a list of *Bradyrhizobium* genome accessions (primarily type strains of validly published species, selected based on the List of Prokaryotic names with Standing in Nomenclature taxonomy), which were then retrieved from GenBank using the GenFlow [12] pipeline. GenFlow automates genome downloading via NCBI command-line tools. All genomes were renamed for clarity with EDirect tool and processed with Anvi'o v7.1 [13]. To construct the 16S rRNA gene tree, we extracted 16S rRNA sequences from genomes using the `anvi-get-sequences-for-hmm-hits` function in Anvi'o v7.1 [13], which employs HMMER v3.3.2 [14] to identify genes. For the following strains: *Bradyrhizobium americanum* CMVU44^T (KU991833.1), *B. ferriligni* CCBAU 51502^T (KX683400.1), *B. ganzhouense* RITF806^T (JQ796661.2), *B. ingae* BR 10250^T (KF927043.1), *B. kavangense* 14-3^T (KP899562.1), *B. namibiense* 5-10^T (KX661401.2), *B. ripae* WR4^T (MF593081.1) and *B. subterraneum* 58 2-1^T (KP308152.1), genome accession numbers were not available in the NCBI database. Consequently, we retrieved their 16S rRNA gene sequences and aligned them with the 16S sequences extracted from the available *Bradyrhizobium* genomes. Sequence alignment was performed with MAFFT v7.397 [15], and a maximum likelihood phylogeny was inferred with IQ-TREE v2.2.0 [16], using the GTR+G+I model and 1,000 bootstrap replicates.

For phylogenomic analyses, coding sequences were annotated using Prodigal v2.6.3 [17], and Single Copy Core Genes across genomes were identified and concatenated with Anvi'o v7.1 [13], applying a geometric and functional index threshold of 0.8, to ensure the inclusion of gene clusters that are broadly conserved and functionally consistent. The geometric homogeneity index quantifies the uniformity of gene distribution across genomes, whereas the functional homogeneity index evaluates the consistency of gene function within clusters. Only gene clusters that met these homogeneity criteria were selected for downstream analyses. These sequences were then aligned using MAFFT v7.397 [15] and generated robust phylogenomic trees using FastTree v2.1.11 [18] with the JTT+CAT model and 1,000 bootstrap resampling. For both trees, visualization was performed using Interactive Tree of Life (iTOL) v6 [19].

To further evaluate the genomic relationships of the studied isolates in comparison to their closest relatives, the average nucleotide identity (ANI) value was calculated using the OrthoANIu [20] algorithm implemented in EzBioCloud [21], and digital DNA–DNA hybridization (dDDH) was determined using the Genome-to-Genome Distance Calculator 3.0 algorithm, based on formula 2 (*d*₄), which calculates the proportion of identities over high-scoring segment pairs [22]. In Table 2, we show the most closely related species to each isolate based on ANI values, whereas Table 3 lists the corresponding dDDH values.

The four *Bradyrhizobium* isolates from this work are positioned in distinct clades relative to other *Bradyrhizobium* species in the phylogenetic trees, each supported by high confidence bootstrap values (Figs 1 and 2). In addition, the isolates exhibited ANI and dDDH values below the commonly accepted species boundaries of 95–96% for ANI and 70% for dDDH when compared with their closest described *Bradyrhizobium* type strains (Tables 2 and 3), supporting their classification as putative new species. Furthermore, the presence of diverse *nod* and *nif* genes suggests that these isolates have various capacities for nodulation and nitrogen fixation, possibly influenced by their host plants and environmental conditions. By characterizing isolates from underexplored tropical regions and host species, our study contributes to a broader understanding of the genomic diversity and evolutionary adaptations of *Bradyrhizobium*. Future phenotypic analyses will be crucial to confirm their species status and elucidate their ecological and biological roles.

Funding information

This study was funded by the University of Costa Rica (projects B9-204 and C0-524) and by the Ministerio de Ciencia, Innovación, Tecnología y Telecomunicaciones de Costa Rica (MICITT)-Consejo Nacional para Investigaciones Científicas y Tecnológicas (CONICIT) (project FI-036B-19).

Acknowledgement

We thank the Colaboratorio Nacional de Computación Avanzada (CNCA) at the Centro Nacional de Alta Tecnología (CeNAT), Costa Rica, for providing access to the Kabré supercomputer.

Author contributions

Conceptualization: R.F.V., S.J.A., V.L.A., K.R.J. and B.M.G.; Methodology: R.F.V., S.J.A., V.L.A. and B.M.G.; Formal analysis: R.F.V., S.J.A. and V.L.A.; Investigation: R.F.V., S.J.A. and V.L.A.; Writing – original draft preparation: R.F.V., S.J.A. and V.L.A.; Supervision: K.R.J. and B.M.G.; Visualization: R.F.V., S.J.A., V.L.A. and B.M.G.; Funding acquisition: K.R.J.; Writing – review and editing: R.F.V., S.J.A., V.L.A., K.R.J. and B.M.G. All authors have read and agreed to the published version of the manuscript.

Conflicts of interest

The authors declare that there are no conflicts of interest.

References

1. Helene LCF, Klepa MS, O'Hara G, Hungria M. *Bradyrhizobium archetypum* sp. nov., *Bradyrhizobium australiense* sp. nov. and *Bradyrhizobium murdochi* sp. nov., isolated from nodules of legumes indigenous to Western Australia. *Int J Syst Evol Microbiol* 2020;70:4623–4636.
2. Tapia-García EY, Hernández-Trejo V, Guevara-Luna J, Rojas-Rojas FU, Arroyo-Herrera I, et al. Plant growth-promoting bacteria isolated from wild legume nodules and nodules of *Phaseolus vulgaris* L. trap plants in central and southern Mexico. *Microbiol Res* 2020;239:126522.
3. Hungria M, Menna P, Delamuta JRM. Bradyrhizobium, the ancestor of all rhizobia: phylogeny of housekeeping and nitrogen-fixation genes. In: *Biological Nitrogen Fixation*. John Wiley & Sons, Ltd. pp. 191–202.
4. Mendoza-Suárez M, Andersen SU, Poole PS, Sánchez-Cañizares C. Competition, nodule occupancy, and persistence of inoculant strains: key factors in the Rhizobium-legume symbioses. *Front Plant Sci* 2021;12:690567.
5. Rojas-Jimenez K, Morera-Huertas J, de Bedout-Mora M, Loria-Vinueza B, Zuniga-Orozco A, et al. Exploring symbiotic legume-rhizobia relationships across tropical species. *Plant Biology* 2024. DOI: 10.1101/2024.08.23.609460.
6. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 2018;34:i884–i890.
7. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012;19:455–477.
8. Li H. Seqtk. GitHub; 2025. <https://github.com/lh3/seqtk> [accessed 9 January 2025].
9. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 2013;29:1072–1075.
10. Chklovski A, Parks DH, Woodcroft BJ, Tyson GW. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. *Nat Methods* 2023;20:1203–1212.
11. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol Biol Evol* 2021;38:5825–5829.
12. Mendoza B. GenFlow: a Bioinformatic Workflow for Phylogenomic Analysis; 2024. <https://www.protocols.io/view/genflow-a-bioinformatic-workflow-for-phylogenomic-dnax5afn> [accessed 9 January 2024].
13. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, et al. Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* 2015;3:e1319.
14. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* 2011;39:W29–37.
15. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast fourier transform. *Nucleic Acids Res* 2002;30:3059–3066.
16. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 2015;32:268–274.
17. Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 2010;11:119.
18. Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* 2009;26:1641–1650.
19. Letunic I, Bork P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res* 2024;52:W78–W82.
20. Lee I, Ouk Kim Y, Park S-C, Chun J. OrthoANI: an improved algorithm and software for calculating average nucleotide identity. *Int J Syst Evol Microbiol* 2016;66:1100–1103.
21. Yoon S-H, Ha S-M, Kwon S, Lim J, Kim Y, et al. Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *Int J Syst Evol Microbiol* 2017;67:1613–1617.
22. Meier-Kolthoff JP, Göker M. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nat Commun* 2019;10:2182.

The Microbiology Society is a membership charity and not-for-profit publisher.

Your submissions to our titles support the community – ensuring that we continue to provide events, grants and professional development for microbiologists at all career stages.

Find out more and submit your article at microbiologyresearch.org