Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Comparative genomics of three Bangladeshi rhizobia strains reveals species-specific genomic variations and functional traits

Nazmul Alam Khan
Bangladesh Institute of Nuclear Agriculture (BINA)
Harun-or Rashid
Bangladesh Institute of Nuclear Agriculture (BINA)
Md. Imdadul Hoque
apu.pstu@gmail.com

Bangladesh Institute of Nuclear Agriculture (BINA)

Research Article

Keywords: Rhizobium, comparative genomics, pan-genome, ANI, nodulation, nitrogen fixation, PGPR, bioinoculant

Posted Date: August 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10571885/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Rhizobium species are important symbiotic nitrogen fixers and potential bioinoculants, yet the comparative genomic characteristics of *Rhizobium bangladeshense* BLR175, *Rhizobium binae* BLR195, and *Rhizobium lentis* BLR27 remain insufficiently characterized. This study applied a standardized workflow comprising Prokka annotation, dRep-based identity analysis, Roary pan-genome reconstruction, functional screening, and Proksee visualization. Genome sizes ranged from 5.96 to 6.81 Mb, with 7,287–7,884 predicted coding sequences. *R. lentis* exhibited an N50 of 5,067 bp and the highest hypothetical-protein fraction among the three genomes. Pairwise average nucleotide identity values of 88.47–89.84% and assignment to distinct secondary clusters confirmed substantial genomic divergence among the three strains. The pan-genome contained 18,646 gene clusters, including 1,336 core and 17,310 accessory clusters. All genomes carried major genes associated with nodulation, nitrogen fixation, microaerobic respiration, surface-polysaccharide biosynthesis, dicarboxylate transport, and cytochrome-c maturation. BLR175 showed the broadest inventories of genes associated with metal homeostasis, siderophore production, and colonization; BLR195 was enriched in phosphate-utilization genes and uniquely carried *merA*; and BLR27 exhibited the most extensive phytohormone and stress-mitigation-related gene profile. These findings identify three genomically distinct candidate bioinoculants with complementary functional potential.

Introduction

Rhizobium species are Gram-negative soil bacteria widely recognized for their ability to establish symbiotic associations with leguminous plants and contribute to biological nitrogen fixation. During this interaction, rhizobial cells colonize plant roots, initiate nodule formation, and differentiate into bacteroids that convert atmospheric nitrogen into biologically available forms. This complex process depends on the coordinated function of genes involved in Nod-factor synthesis, nitrogenase activity, microaerobic respiration, surface-polysaccharide production, nutrient transport, motility, and host-root colonization [1]. Therefore, genome-level characterization of Rhizobium species is essential for understanding their symbiotic efficiency, ecological adaptation, evolutionary relationships, and potential agricultural applications [2].

Rhizobial inoculants are increasingly important in sustainable agriculture because they can improve legume productivity, reduce dependence on synthetic nitrogen fertilizers, and support soil fertility [3]. In addition to nitrogen fixation, many Rhizobium strains possess genomic traits associated with phytohormone metabolism, siderophore-mediated iron acquisition, phosphate utilization, rhizosphere colonization, biocontrol, heavy-metal tolerance, and antimicrobial resistance [4]. These traits may enable rhizobia to survive under diverse environmental conditions and promote plant growth under nutrient limitation or abiotic stress. However, the presence of a gene in a bacterial genome indicates predicted functional potential and does not necessarily confirm the corresponding phenotype without experimental validation [5].

Comparative genomics provides an effective framework for examining both conserved and variable characteristics among bacterial genomes [6]. Genome annotation enables the identification of protein-coding sequences, transfer RNA genes, ribosomal RNA genes, and putative functional products [7], whereas average nucleotide identity analysis provides a quantitative measure of genome-wide relatedness [8]. Pan-genome analysis further distinguishes conserved core genes from accessory and genome-specific genes, thereby revealing the extent of genetic diversity and possible niche specialization among related organisms [9]. In rhizobia, important symbiosis-associated determinants include the nod, nol, and noe gene families involved in nodulation [10]; the nif genes required for nitrogenase synthesis and cofactor formation [11]; and the fix and cco genes associated with respiration under the low-oxygen conditions of root nodules [12]. Other gene families, including exo, pss, rkp, ndv, chv, dct, and ccm, contribute to extracellular-polysaccharide production, host infection, bacteroid carbon metabolism, and cytochrome c maturation [13,14].

The genomes of *Rhizobium binae* BLR195, *Rhizobium bangladeshense* BLR175, and *Rhizobium lentis* BLR27 are available in public repository databases, but their genomic structure, annotation depth, nucleotide-level relatedness, pan-genome composition, and functional differentiation have not been sufficiently evaluated using a single standardized analytical framework. Comparative information is particularly limited regarding the distribution of their nodulation, nitrogen-fixation, infection-support, plant growth-promoting, metal-homeostasis, phosphate-acquisition, siderophore-associated, rhizosphere-colonization, and multidrug-efflux genes. Moreover, their assembly characteristics have not been adequately contextualized against accession-verified Rhizobium genomes from different geographical regions.

To address these gaps, the present study performed an integrated comparative genomic analysis of R. binae BLR195, R. bangladeshense BLR175, and R. lentis BLR27. A standardized bioinformatics workflow was applied for genome annotation, assessment of assembly and

annotation characteristics, average nucleotide identity analysis, pan-genome analysis, functional gene screening, and circular genome visualization. Symbiosis-associated gene families and genes related to plant growth promotion and environmental adaptation were systematically compared among the three genomes.

Materials and Methods

Genome datasets and bioinformatics workflow

Three genomes *Rhizobium binae* BLR195 (GCA_017357225.1), *R. bangladeshense* BLR175 (GCA_017357245.1), and *R. lentis* BLR27 (GCA_017352135.1) were obtained from the NCBI Genome database [15] and originally described by Rashid et al. (2015)[16]. The workflow included Prokka annotation [17], dRep based ANI analysis [18], Roary pan-genome analysis [19], and Proksee visualization [20].

Genome annotation using Prokka

The three genome assemblies were independently annotated using Prokka [17] in bacterial mode through Galaxy [21]. Prokka predicted CDS, tRNA, rRNA, and other genomic features. GFF3 files were used for pan-genome analysis, while TSV and TXT outputs provided annotation statistics; LOG and ERR files were checked to confirm successful processing.

Assessment of genome assembly and annotation characteristics

Genome size and contig number were determined from the FASTA files, while CDS, gene, rRNA, and tRNA counts were obtained from Prokka outputs [17]. CDS and contig densities were calculated per megabase, and mean and median CDS lengths were derived from GFF3 coordinates. Approximate coding density was estimated as the summed CDS length divided by total genome length and expressed as a percentage.

Evaluation of annotation depth

Annotation depth was assessed using Prokka-derived gene symbols [17], Enzyme Commission (EC) numbers, Clusters of Orthologous Groups (COG) assignments, and product descriptions. Coding sequences (CDSs) annotated as hypothetical proteins were counted separately. The proportions of hypothetical, non-hypothetical, gene-symbol-assigned, EC-annotated, and COG-annotated CDSs were calculated relative to the total number of coding sequences.

Identity and genome clustering

Genomic relatedness among the three *Rhizobium* genomes was assessed using dRep through Galaxy [18,21]. Primary clustering was performed by average-linkage clustering at a distance threshold of 0.10, followed by secondary comparison using average nucleotide identity based on MUMmer with filtering (ANImf) [22]. Pairwise average nucleotide identity (ANI) [23], alignment coverage, reference coverage, and query coverage were extracted, and reciprocal ANI values were averaged to generate the similarity matrix.

Pan-genome analysis

Pan-genome analysis was performed using Roary [19] through Galaxy [21] with Prokka-generated [17] General Feature Format version 3 files. Homologous coding sequences were clustered into orthologous groups, and a gene presence–absence matrix was generated. Gene clusters were classified as core (99-100%), soft core (95-<99%), shell (15-<95%), or cloud (>0-<15%) according to their distribution among genomes. Core, accessory, shared, and genome-specific gene clusters were extracted from the Roary outputs.

Functional annotation and screening

Functional categories were identified from Prokka-derived gene symbols and product descriptions [17]. Coding sequences were grouped into transport, transcriptional regulation, redox metabolism, biosynthesis, motility and chemotaxis, cell-envelope functions, and nitrogen-associated processes. Category counts represented predicted annotation hits.F

Identification of symbiosis-associated genes

Symbiosis-related genes were screened from Prokka gene-symbol annotations [17]. Nodulation genes included the *nod*, *nol*, and *noe* families; nitrogen-fixation genes included *nif*, *fix*, and *cco*; and infection-support genes included *exo*, *pss*, *rkp*, *ndv*, and *chv*. The *dct*, *ccm*, *gln/ntr*, and *fla/flg/flh/fli/mot/che* families were used to identify dicarboxylate transport, cytochrome *c* maturation, nitrogen regulation, and

motility or chemotaxis functions, respectively. Prokka numeric suffixes were retained for copy-number analysis but removed for binary presence–absence scoring. Genes lacking a direct symbol match were recorded as not detected rather than definitively absent.

Screening of plant growth-promoting and stress-adaptation traits

A curated Prokka-based [17] gene inventory was used to identify traits associated with heavy-metal tolerance, phytohormone production, biocontrol, siderophore-mediated iron acquisition, phosphate utilization, quorum sensing, rhizosphere colonization, and antimicrobial resistance. Relevant gene symbols and product annotations were grouped into these functional categories, and each predicted coding sequence was counted as an annotation hit. The resulting counts represented genomic potential only and were not considered evidence of experimentally confirmed phenotypes.

Reference genome comparison

The three genomes were compared with four *Rhizobium* reference genomes retrieved from the National Center for Biotechnology Information (NCBI): *R. leguminosarum* I10 from India (GCA_051904995.1), *R. favelukesii* T136 from Canada (GCA_021044665.1), *Rhizobium* sp. SARCC-755 from South Africa (GCF_008370325.1), and *R. etli* CFN42 from Mexico (GCA_000092045.1/GCF_000092045.1) [24]. Genome size, GC content, sequence count, assembly status, and replicon organization were compared.

Data analysis and graphical presentation

Results were summarized using descriptive statistics, including counts, means, medians, percentages, densities, and gene presence-absence patterns. All bar charts and heatmaps, including the pairwise average nucleotide identity matrix and symbiosis-associated gene-family heatmap, were generated using GraphPad Prism version 8. Pan-genome clusters were presented according to genome-specific, pairwise-shared, and core distribution patterns.

Results

Genome assembly and annotation summary

The three genomes ranged from 5.960 Mb to 6.806 Mb. *R. binae* was the largest genome and contained the highest CDS count, while *R. bangladeshenes* had the fewest CDS (Table 1). *R. lentis* had a much higher contig count than the other two genomes. CDS density ranged from 1,164 to 1,308 CDS/Mb, and approximate coding density ranged from 77.43% to 84.62%.

Table 1. Genome annotation and assembly metrics.

Genome	Contigs	Genome size (Mb)	CDS	Genes	CDS/Mb	Contigs/Mb	Mean CDS length (bp)	Median CDS length (bp)	Approx. coding density (%)	rRNA	tRNA
R. binae	444	6.81	7,884	8,010	1,158	65.23	718.4	618.0	83.21	3	48
R. bangladeshenes	254	6.26	7,287	7,395	1,164	40.56	727.3	633.0	84.62	3	42
R. lentis	1,566	5.96	7,797	7,891	1,308	262.8	591.9	498.0	77.43	3	39

Annotation depth and hypothetical proteins

Prokka annotation depth varied among the genomes. *R. bangladeshenes* had the lowest hypothetical-protein fraction, while *R. lentis* had the highest (Figure 1). The higher hypothetical-protein rate in *R. lentis* may reflect true novelty, lower annotation database match, or its more contig no assembly. EC-number and COG assignment rates also differed, providing a useful estimate of how much of each genome received functional annotation beyond a generic product name.

dRep clustering and ANI-based relatedness

The dRep clustering output placed the three genomes in one primary cluster but assigned them to different secondary clusters. Pairwise ANI values were below 90%, with the highest mean ANI between *R. lentis* and *R. bangladeshenes* and the lowest mean ANI between *R. binae* and *R. bangladeshenes* (Figure 2). Because all pairwise ANI values are far below the commonly used 95-96% species-level reference boundary, the genomes represent clearly divergent Rhizobium lineages rather than near-identical strains.

Roary pan-genome composition

Roary identified 18,646 total pan-genome gene clusters across the three assemblies. Only 1,336 clusters were present in all three genomes, whereas 17,310 clusters were classified as shell/accessory genes. With only three genomes, soft-core and cloud categories were zero because the presence thresholds collapse into core or shell patterns. The core fraction was 7.17% of the pan-genome, indicating extensive gene-content diversity among the three genomes (Table 2).

Table 2. Roary pan-genome summary statistics.

Category	Definition	Gene clusters	Percentage of pan-genome (%)
Core genes	(99% ≤ strains ≤ 100%)	1,336	7.17
Soft core genes	(95% ≤ strains < 99%)	0	0.00
Shell genes	(15% ≤ strains < 95%)	17,310	92.83
Cloud genes	(0% ≤ strains < 15%)	0	0.00
Total genes	(0% ≤ strains ≤ 100%)	18,646	100.0

Presence-pattern analysis showed that genome-specific clusters dominated the pan-genome. *R. binae* had the highest number of genome-specific clusters, followed by *R. lentis* and *R. bangladeshenes*. The largest two-genome shared fraction was between *R. lentis* and *R. bangladeshenes*, consistent with their relatively higher ANI value compared with the other pairs.

Functional category trends

A scan of large numbers of transporters, transcriptional regulators, oxidoreductases/ dehydrogenases, and transferase/ ligase/ synthase proteins. These categories are typical of metabolically flexible soil bacteria. Cell-envelope/polysaccharide-related annotations were also detected in all genomes, supporting the presence of surface-structure functions relevant to rhizosphere survival and symbiotic infection.

Symbiosis-associated gene-family counts

Gene-family screening detected multiple copies or variants of nodulation, nitrogenase, fixation, surface polysaccharide, dicarboxylate transport, cytochrome c maturation, nitrogen regulation, and motility/chemotaxis genes. *R. binae* and *R. bangladeshenes* had more *nod/nol/noe* and *nif* gene-symbol matches than *R. lentis*, whereas R. lentis had relatively high *fix/cco* and motility/chemotaxis counts (Figure 4).

Key nodulation, nitrogen-fixation, and infection-support genes

A compact presence/absence screen of selected genes showed that all three genomes carried nodA and nodM, supporting the presence of Nod-factor-related gene candidates. *nodF* was detected in *R. binae* and *R. bangladeshenes*, while *nodO* was detected in *R. bangladeshenes* and *R. lentis*.

The nitrogenase structural genes *nifH, nifD*, and *nifK* were detected in all three genomes, along with *nifB. nifA* was detected in *R. binae* and *R. bangladeshenes* but not in *R. lentis* (Table 3). Among low-oxygen fixation genes, *fixK* and *fixN* were detected in all three genomes, while *fixL, fixJ, fixX*, *fixA*, and *fixB* showed variable detection. Infection and surface-support genes *ndvA, ndvB, rkpK*, and *chvE* were detected across all three genomes, and *dctB/dctD/dctM* were also detected, supporting potential C4-dicarboxylate transport capacity needed for bacteroid metabolism.

Table 3. Presence/absence of selected symbiosis-associated genes gene screening.

Functional group	Gene	R. binae	R. bangladeshenes	R. lentis	Function of the gene
Nod factor	*nodA*	+	+	+	Acyltransferase, Nod-factor synthesis [25]
Nod factor	*nodF*	+	+	-	Acyl-ACP, Nod-factor acylation [25]
Nod factor	*nodM*	+	+	+	Glucosamine synthase, Nod precursor [26]
Nod factor	*nodO*	-	+	+	Ca^{2+}-binding signalling pore [27]
Nitrogenase / cofactor	*nifH*	+	+	+	Fe-protein of nitrogenase [28]
Nitrogenase / cofactor	*nifD*	+	+	+	MoFe-protein α-subunit [25]
Nitrogenase / cofactor	*nifK*	+	+	+	MoFe-protein β-subunit [26]
Nitrogenase / cofactor	*nifA*	+	+	-	Nif transcriptional activator [29]
Nitrogenase / cofactor	*nifB*	+	+	+	FeMo-cofactor biosynthesis [30]
Microaerobic fixation	*fixK*	+	+	+	FNR-like transcription factor [12]
Microaerobic fixation	*fixL*	+	-	-	O_2-sensing histidine kinase [12]
Microaerobic fixation	*fixJ*	+	-	+	Two-component response regulator [12]
Microaerobic fixation	*fixN*	+	+	+	cbb3 cytochrome oxidase subunit [12]
Microaerobic fixation	*fixX*	+	-	+	Ferredoxin-like electron carrier [12]
Surface / infection support	*pssM*	+	+	-	Exopolysaccharide biosynthesis [31]
Surface / infection support	*ndvA*	+	+	+	Cyclic β-glucan exporter [31]
Surface / infection support	*ndvB*	+	+	+	Cyclic β-glucan synthase [31]
Surface / infection support	*rkpK*	+	+	+	K-antigen capsular biosynthesis [31]
Surface / infection support	*chvE*	+	+	+	Sugar-binding chemotaxis signal [31]
Bacteroid carbon supply	*dctA*	+	+	-	C_4-dicarboxylate permease [12]
Bacteroid carbon supply	*dctB*	+	+	+	C_4-dicarboxylate sensor kinase [12]
Bacteroid carbon supply	*dctD*	+	+	+	C_4-dicarboxylate response regulator [12]
Bacteroid carbon supply	*dctM*	+	+	+	TRAP transporter subunit [12]
Cytochrome c maturation	*ccmA*	+	+	+	Heme ABC transporter ATPase [12]
Cytochrome c maturation	*ccmB*	+	+	+	Heme ABC transporter membrane [12]
Cytochrome c maturation	*ccmC*	+	+	+	Inner-membrane heme relay [12]
Cytochrome c maturation	*ccmF*	+	+	+	Apocytochrome c heme lyase [12]
Cytochrome c maturation	*ccmH*	+	-	+	Heme lyase maturation factor [12]

Note: “+” = detected; “-” = not detected

Species specific plant growth-promoting and stress-adaptation traits

A total of 397 category-assigned Prokka CDS annotation hits were included in the comparative inventory. R. bangladeshenes had the highest overall count (139), followed by *R. binae* (131) and *R. lentis* (127) (Figure 6). The three genomes displayed broadly similar

antimicrobial/biocontrol candidate counts, but differed markedly in metal homeostasis, siderophore-associated functions, phosphate acquisition, phytohormone-related traits, and colonization-associated systems.

Comparative genomic analysis of plant growth-promoting and stress-adaptation traits

Comparative genomic analysis revealed substantial variation among *R. binae*, *R. bangladeshense*, and *R. lentis* in functional traits associated with metal tolerance, plant growth promotion, rhizosphere adaptation, and antimicrobial resistance. *R. bangladeshense* showed the highest number of metal-associated annotations (32), followed by *R. lentis* (28) and *R. binae* (24) (Supplementary Table 6). All three genomes possessed arsenic resistance-related genes and multiple copper- and divalent-cation transport systems. *R. bangladeshense* exhibited the broadest metal homeostasis repertoire, including *ars, czc, znt, cnr, chrB1, srpC*, *silP*, and multiple *cop/cus/actP* genes. In contrast, *R. binae* uniquely contained *merA*, suggesting potential mercury reduction capability, whereas *R. lentis* possessed additional chromate- and cadmium-associated determinants (Supplementary Table 7).

Phytohormone-associated traits varied among the genomes, with *R. lentis* containing the largest inventory (8 annotations), followed by *R. binae* (5) and *R. bangladeshense* (4). Both *R. bangladeshense* and *R. lentis* carried *ipdC*, suggesting potential involvement in indole-3-acetic acid biosynthesis, while *R. lentis* additionally contained multiple *acdS* copies, indicating a stronger potential for ACC deaminase-mediated stress mitigation. No complete *iaaM-iaaH* or gibberellin biosynthetic pathway was identified (Supplementary Table 7).

R. bangladeshense displayed the strongest siderophore-associated profile, with 19 iron acquisition-related annotations, including *dhbF, iucD, mbtH*, and multiple *fhu* transport genes. *R. binae* showed the highest phosphate-related inventory (20 annotations), including *gcd* and *phn* genes involved in phosphate utilization. Furthermore, *R. bangladeshenes* possessed the largest quorum sensing and colonization-related gene repertoire (20 annotations), including *luxR, rpaR, vjbR*, and extracellular polysaccharide-associated genes. Multidrug-efflux-associated annotations were abundant in all genomes, with *R. binae* containing the highest number (51), followed by *R. bangladeshense* and *R. lentis* (47 each). Overall, *R. bangladeshenes* demonstrated the broadest genomic potential for stress adaptation and rhizosphere colonization, while *R. binae* and *R. lentis* exhibited specialized functional advantages in mercury resistance and plant stress-related traits, respectively (Supplementary Table 7).

Detailed comparative gene inventory of plant growth-promoting and stress-adaptation traits

Accession-verified international structural context

The local assemblies were compared structurally with four external genomes meeting strict accession and provenance criteria (Table 7). The mean local assembly length was 6,343,069 bp, whereas the external-panel mean was 7,041,818 bp; the local mean was 9.92% lower. Local GC contents fell within the verified external range of 59.5-61.5%. Assembly continuity was markedly lower for all local genomes than for the Indian I10 and South African SARCC-755 draft assemblies, and the Canadian T136 and Mexican CFN42T genomes were complete multipartite assemblies.

Discussion

The integrated comparative-genomic framework Prokka annotation, dRep clustering, Roary pan-genome reconstruction and Proksee, executed against accession-verified international references yields a coherent portrait of three divergent Bangladeshi Rhizobium lineages (*R. binae BLR195, R. bangladeshense BLR175, R. lentis BLR27*), with interpretive caveats consolidated in the closing limitations.

The 5.96–6.81 Mb genome-size window sits within the broader Rhizobium landscape Rhizobiales genomes can scale up to 9 Mbp with several mega-sized plasmids [32] and aligns with the 8.01 Mb Tunisian *Rhizobium acaciae* sp. nov. type strain 1AS11T [33], placing all three isolates within the body of the genus-size distribution. The Contigs/Mb ratio of *R. lentis* (262.8) signals large contig number, contrasting with the high-quality guar-nodulating *Rhizobium* sp. RCAM05973 draft (6,937,221 bp in 2 contigs) [34]. CDS density (1,158-1,308 CDS/Mb) and coding density (77.43–84.62%) are typical for soil Rhizobiaceae; *R. lentis*'s 77.43% reflects compressed average CDS length (591.9 bp versus 718.4-727.3 bp) driven non-coding gaps. Prokka depth metrics (gene-symbol CDS 40.68–48.50%, EC 26.05–29.94%, COG 25.77–29.96%, hypothetical fraction 48.62–56.48%) are typical for automated prokaryotic annotation-unannotated CDS can reach 70% in poorly characterised genomes but only 27% in the curated *Mycobacterium tuberculosis* H37Rv proteome [35] and a comparable 16% in *Shigella flexneri* (674 hypothetical proteins among 4,362 total) [36] and *R. lentis*'s 56.48% likely reflects biological novelty, weak database representation, and downsampling of fragmented ORFs.

Pairwise ANI values of 88.47–89.84% lie well below the 95–96% ANI / 70% dDDH species-delimitation consensus, reaffirmed across 270,400 pairwise comparisons of 520 Rhizobiales type-strains [37], and well below the 96% ANI gap that naturally demarcates 18 distinct

genospecies (plus 7 unique strains) in the *Rhizobium leguminosarum* complex [38]. *Rhizobium* is notoriously paraphyletic, and a contemporary Rhizobiaceae reclassification proposed a pairwise core-proteome average amino acid identity (cpAAI) threshold of ~ 86% for genus delimitation [39]. The dRep primary-cluster placement at 0.10 reflects average-linkage secondary clustering rather than conspecificity; under ANImf partitioning, each genome is assigned to its own secondary cluster.

The 18,646-cluster pan-genome with only 1,336 core (7.17%) and 17,310 shell (92.83%) clusters and no soft-core or cloud clusters-conforms to the open pan-genome architecture typical of rhizobia. Soft-core and cloud collapse to zero is an inevitable consequence of the n = 3 input under standard Roary thresholds, so the 7.17% core fraction is an upper estimate and the 92.83% accessory fraction an overstatement, because fragmented drafts inflate accessory content. Genome-specific clusters (5,626 *R. binae*, 5,405 *R. lentis*, 4,629 *R. bangladeshense*) align with documented variability of plasmid-borne loci in *Rhizobium*; the largest pairwise-shared fraction between *R. lentis* and *R. bangladeshense* (728 clusters, 3.90%) reflects their higher pairwise ANI (89.84%).

The functional census (transporters 527–575, regulators 432–543, redox enzymes 479–504, biosynthetic enzymes 544–596, motility 41–48, cell-envelope 26–28, nitrogen-associated 17–27) mirrors the canonical profile of soil-resident rhizobia, with 500-hit transport/regulatory inventories reflecting the metabolic flexibility required for heterogeneous rhizosphere colonisation.

Universal detection of *nodA*, *nodM*, *nifH*, *nifD*, *nifK*, *nifB*, *fixK*, *fixN*, *ndvA*, *ndvB*, *rkpK*, *chvE*, *dctB*, *dctD*, *dctM* and the *ccmA/B/C/F* operon confirms foundational Nod-factor synthesis, complete nitrogenase machinery, microaerobic respiration, surface-polysaccharide support, C4-dicarboxylate transport and cytochrome c biogenesis. The Mo-Fe pathway depends on *nifB*-driven Fe-S formation and the *nifEN* scaffold, with *nif* expression linked via *nifA*, *nifL* and *glnK*. The apparent absence of *nodB*, *nodC* and *nodD* as direct Prokka hits does not reflect true biological absence: 41 bean-nodulating Mexican *Rhizobium* isolates surveyed across acid and alkaline soils carry only subsets of the canonical *nod* cluster (*nodC* and *nifH* absent from *R. grahamii* and *Rhizobium radiobacter*) [40], and these loci may be present under alternative symbols (e.g., *nolL*, *nodD1*, *nodD2*) or fragmented. Higher *fix/cco* and motility/chemotaxis counts in *R. lentis* balanced against higher *nod/nol/noe* (9–10) and *nif* (9–11) counts in *R. binae* and *R. bangladeshense* are consistent with plasmid-based partitioning (fix, dct and phb chromosomal; nod and nif plasmid-borne on pSym) reported across *Rhizobium*.

R. bangladeshense BLR175 carries the broadest metal-homeostasis inventory (32 hits: *czc*, *znt*, *cnr*, *chrB1*, *srpC*, *silP*, *cop/cus/actP*), consistent with the prevalence of metal-resistance determinants in rhizobia from contaminated soils [41]. Unique *merA* in *R. binae* requires guarded interpretation: independent *merA* copies are widespread across alphaproteobacteria and hypertolerance requires a complete Mer operon. *R. lentis* BLR27 carries the largest phytohormone inventory (8 hits: multiple *acdS*, *ipdC* and a LOG-like cytokinin activator); ACC deaminase acts as a metabolic sink for plant ACC, reducing ethylene-mediated stress inhibition under abiotic stresses including salinity and drought, in line with its designation as a "stress modulator" [42]. *R. bangladeshense* shows the strongest siderophore profile (19 hits: *dhbF*, *iucD*, *mbtH*, expanded *fhu*); Rhizobiales iron homeostasis is regulated by the lineage-specific regulator RirA, whose deletion in *Sinorhizobium* derepresses a heterologous *fprA-asb-fat* petrobactin polycistron, causing iron overload in nodules and abolishing nitrogen fixation [43]. *R. binae* BLR195 carries the largest p hosphate-utilisation inventory (20 hits: *gcd*, near-complete *phn* phosphonate cluster); Proteobacteria, including Rhizobiaceae, dominate *gcd*-mediated gluconic-acid solubilisation in soil metagenomes. Multidrug-efflux counts (47–51 hits) are abundant and similar across all three, consistent with the intrinsic efflux arsenal carried by soil Proteobacteria and with Burkholderiaceae-dominated soil resistomes in which multidrug efflux pump genes are the discriminatory ARG type regardless of antibiotic selection [44].

The mean local assembly length of 6,343,069 bp sits 9.92% below the accession-verified external mean of 7,041,818 bp, reflecting the multipartite replicon structure typical of Rhizobiales, in which 10% of bacterial genomes are split between multiple large DNA molecules[32] exemplified by the tripartite *Sinorhizobium meliloti* 1021 architecture (3.65 Mbp chromosome + 1.35 Mbp pSymA + 1.68 Mbp pSymB). The Indian I10 (7 contigs) and South African SARCC-755 (36 contigs) drafts retain higher contiguity, while the Canadian *R. favelukesii* T136 (chromosome 4.2 Mb plus 4 plasmids) [45]and Mexican *R. etli* CFN42 (circular chromosome plus six plasmids)[46] are complete multipartite genomes; resolving such architectures is essential for capturing plasmid-encoded symbiotic regions, as demonstrated by the closed *R. leguminosarum* sv. *trifolii* TA1 and CC275e commercial clover-inoculant genomes [47].

Comparative-genomic evidence supports *R. bangladeshense* BLR175, *R. binae* BLR195 and *R. lentis* BLR27 originally proposed as three distinct species by Rashid et al. (2015) as three distinct candidate bioinoculants whose complementary trait profiles broad metal homeostasis in BLR175, potential Hg-reductase activity and strong phosphate utilisation in BLR195, and phytohormone plus stress-mitigation inventory in BLR27 could be exploited as a consortium. This mirrors the predictive inoculant-strain pipeline developed for Indian pigeon-pea endosymbionts, in which *Bradyrhizobium yuanmingense* and *Ensifer alkalisoli* strains were ranked by combined

nodulation efficiency and root-colonisation competitiveness [48], and the broader finding that inoculant selection for nodulation-deficient legumes must weigh competitiveness in soil, not just N_2-fixing potential [49].

Conclusion

This study demonstrates that *R. bangladeshense* BLR175, *R. binae* BLR195, and *R. lentis* BLR27 are genomically distinct *Rhizobium* lineages with conserved symbiotic potential but contrasting accessory functions. Pairwise ANI values below 90% and separate dRep secondary clusters confirm substantial divergence, while the small core genome and extensive accessory fraction indicate marked gene-content variability. All three genomes contained key nodulation, nitrogenase, microaerobic-respiration, surface-polysaccharide, dicarboxylate-transport, and cytochrome-*c*-maturation genes, supporting their predicted capacity for legume association and nitrogen fixation. Their functional profiles were complementary: BLR175 was enriched in metal-homeostasis, siderophore, and colonization genes; BLR195 showed the strongest phosphate-utilization potential and carried *merA*; and BLR27 possessed the largest phytohormone and stress-mitigation inventory. These differences suggest potential use as individual inoculants .

Declarations

Author Contribution

NAK performed the analyses, data interpretation, visualization, and drafted the manuscript. HOR contributed to scientific interpretation, validation, and manuscript revision. MIH supervised the study, provided guidance, data interpretation, visualization and critically revised the manuscript. All authors read and approved the final manuscript.

References

1. Yang J, Lan L, Jin Y, Yu N, Wang D, Wang E. Mechanisms underlying legume–rhizobium symbioses. *J Integr Plant Biol*. 2022;64(2):244-267. doi:10.1111/jipb.13207
2. Goyal RK, Mattoo AK, Schmidt MA. Rhizobial–Host Interactions and Symbiotic Nitrogen Fixation in Legume Crops Toward Agriculture Sustainability. *Front Microbiol*. 2021;12. doi:10.3389/fmicb.2021.669404
3. Abd-Alla MH, Al-Amri SM, El-Enany AWE. Enhancing Rhizobium–Legume Symbiosis and Reducing Nitrogen Fertilizer Use Are Potential Options for Mitigating Climate Change. *Agriculture*. 2023;13(11):2092. doi:10.3390/agriculture13112092
4. Andrade de LA, Santos CHB, Frezarin ET, Sales LR, Rigobelo EC. Plant Growth-Promoting Rhizobacteria for Sustainable Agricultural Production. *Microorganisms*. 2023;11(4):1088. doi:10.3390/microorganisms11041088
5. Jaiswal SK, Mohammed M, Ibny FYI, Dakora FD. Rhizobia as a Source of Plant Growth-Promoting Molecules: Potential Applications and Possible Operational Mechanisms. *Front Sustain Food Syst*. 2021;4. doi:10.3389/fsufs.2020.619676
6. Zhong C, Hu G, Hu C, Xu C, Zhang Z, Ning K. Comparative genomics analysis reveals genetic characteristics and nitrogen fixation profile of Bradyrhizobium. *iScience*. 2024;27(2):108948. doi:10.1016/j.isci.2024.108948
7. Li W, O'Neill KR, Haft DH, et al. RefSeq: expanding the Prokaryotic Genome Annotation Pipeline reach with protein family model curation. *Nucleic Acids Res*. 2021;49(D1):D1020-D1028. doi:10.1093/nar/gkaa1105
8. Olm MR, Crits-Christoph A, Diamond S, Lavy A, Matheus Carnevali PB, Banfield JF. Consistent Metagenome-Derived Metrics Verify and Delineate Bacterial Species Boundaries. *mSystems*. 2020;5(1). doi:10.1128/msystems.00731-19
9. Tonkin-Hill G, MacAlasdair N, Ruis C, et al. Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biol*. 2020;21(1):180. doi:10.1186/s13059-020-02090-4
10. Provorov NA, Andronov EE, Kimeklis AK, Onishchuk OP, Igolkina AA, Karasev ES. Microevolution, speciation and macroevolution in rhizobia: Genomic mechanisms and selective patterns. *Front Plant Sci*. 2022;13. doi:10.3389/fpls.2022.1026943
11. Wekesa CS, Furch ACU, Oelmüller R. Isolation and Characterization of High-Efficiency Rhizobia From Western Kenya Nodulating With Common Bean. *Front Microbiol*. 2021;12. doi:10.3389/fmicb.2021.697567
12. Wang S, Meade A, Lam HM, Luo H. Evolutionary Timeline and Genomic Plasticity Underlying the Lifestyle Diversity in Rhizobiales. *mSystems*. 2020;5(4). doi:10.1128/mSystems.00438-20
13. Acosta-Jurado S, Fuentes-Romero F, Ruiz-Sainz JE, Janczarek M, Vinardell JM. Rhizobial Exopolysaccharides: Genetic Regulation of Their Synthesis and Relevance in Symbiosis with Legumes. *Int J Mol Sci*. 2021;22(12):6233. doi:10.3390/ijms22126233

14. Tong W, Li X, Wang E, et al. Genomic insight into the origins and evolution of symbiosis genes in Phaseolus vulgaris microsymbionts. *BMC Genomics*. 2020;21(1):186. doi:10.1186/s12864-020-6578-0
15. Sayers EW, Bolton EE, Brister JR, et al. Database resources of the national center for biotechnology information. *Nucleic Acids Res*. 2022;50(D1):D20-D26. doi:10.1093/nar/gkab1112
16. Rashid MH or, Young JPW, Everall I, et al. 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. *Int J Syst Evol Microbiol*. 2015;65(Pt_9):3037-3045. doi:10.1099/ijs.0.000373
17. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014;30(14):2068-2069. doi:10.1093/bioinformatics/btu153
18. Olm MR, Brown CT, Brooks B, Banfield JF. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J*. 2017;11(12):2864-2868. doi:10.1038/ismej.2017.126
19. Page AJ, Cummins CA, Hunt M, et al. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics*. 2015;31(22):3691-3693. doi:10.1093/bioinformatics/btv421
20. Grant JR, Enns E, Marinier E, et al. Proksee: in-depth characterization and visualization of bacterial genomes. *Nucleic Acids Res*. 2023;51(W1):W484-W492. doi:10.1093/nar/gkad326
21. Abueg LAL, Afgan E, Allart O, et al. The Galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update. *Nucleic Acids Res*. 2024;52(W1):W83-W94. doi:10.1093/nar/gkae410
22. Lindsey RL, Gladney LM, Huang AD, et al. Rapid identification of enteric bacteria from whole genome sequences using average nucleotide identity metrics. *Front Microbiol*. 2023;14. doi:10.3389/fmicb.2023.1225207
23. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun*. 2018;9(1):5114. doi:10.1038/s41467-018-07641-9
24. O'Leary NA, Wright MW, Brister JR, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res*. 2016;44(D1):D733-D745. doi:10.1093/nar/gkv1189
25. Khairnar M, Alhewairini SS, Kajale SC, et al. Comprehensive genomic analysis of Sinorhizobium meliloti isolates associated with fenugreek (Trigonella foenum-graecum Linn.) from diverse agroclimatic regions of India. *Chemical and Biological Technologies in Agriculture*. 2025;12(1):135. doi:10.1186/s40538-025-00811-0
26. Adedayo AA, Babalola OO. Genomic mechanisms of plant growth-promoting bacteria in the production of leguminous crops. *Front Genet*. 2023;14. doi:10.3389/fgene.2023.1276003
27. Jalmi SK, Sinha AK. Ambiguities of PGPR-Induced Plant Signaling and Stress Management. *Front Microbiol*. 2022;13. doi:10.3389/fmicb.2022.899563
28. Janczarek M, Kozieł M, Adamczyk P, et al. Symbiotic efficiency of Rhizobium leguminosarum sv. trifolii strains originating from the subpolar and temperate climate regions. *Sci Rep*. 2024;14(1):6264. doi:10.1038/s41598-024-56988-1
29. Hakim S, Naqqash T, Nawaz MS, et al. Rhizosphere Engineering With Plant Growth-Promoting Microorganisms for Agriculture and Ecological Sustainability. *Front Sustain Food Syst*. 2021;5. doi:10.3389/fsufs.2021.617157
30. Rivera FN, Guevara-Luna J, Yan J, et al. Rhizobium hidalgonense and Rhizobium redzepovicii as novel faba bean (Vicia faba) microsymbionts in Mexican soils. Preprint posted online October 23, 2023. doi:10.21203/rs.3.rs-3467589/v1
31. 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. doi:10.3389/fpls.2021.690567
32. diCenzo GC, Finan TM. The Divided Bacterial Genome: Structure, Function, and Evolution. *Microbiology and Molecular Biology Reviews*. 2017;81(3). doi:10.1128/MMBR.00019-17
33. Hsouna J, Ilahi H, Han JC, et al. Rhizobium acaciae sp. nov., a new nitrogen-fixing symbiovar isolated from root nodules of Acacia saligna in Tunisia. *Int J Syst Evol Microbiol*. 2023;73(5). doi:10.1099/ijsem.0.005900
34. Guro P, Ulianich P, Belimov A, et al. Draft Genome Sequence of Guar (Cyamopsis tetragonoloba L.) Microsymbiont *Rhizobium* sp. Strain RCAM05973. *Microbiol Resour Announc*. 2023;12(6). doi:10.1128/mra.00071-23
35. Yang Z, Zeng X, Tsui SKW. Investigating function roles of hypothetical proteins encoded by the Mycobacterium tuberculosis H37Rv genome. *BMC Genomics*. 2019;20(1):394. doi:10.1186/s12864-019-5746-6
36. Gazi MdA, Mahmud S, Fahim SM, et al. Functional Prediction of Hypothetical Proteins from Shigella flexneri and Validation of the Predicted Models by Using ROC Curve Analysis. *Genomics Inform*. 2018;16(4):e26. doi:10.5808/GI.2018.16.4.e26

37. Volpiano CG, Sant'Anna FH, Ambrosini A, et al. Genomic Metrics Applied to Rhizobiales (Hyphomicrobiales): Species Reclassification, Identification of Unauthentic Genomes and False Type Strains. *Front Microbiol*. 2021;12. doi:10.3389/fmicb.2021.614957
38. Young JPW, Moeskjær S, Afonin A, et al. Defining the Rhizobium leguminosarum Species Complex. *Genes (Basel)*. 2021;12(1):111. doi:10.3390/genes12010111
39. Kuzmanović N, Fagorzi C, Mengoni A, Lassalle F, diCenzo GC. Taxonomy of Rhizobiaceae revisited: proposal of a new framework for genus delimitation. *Int J Syst Evol Microbiol*. 2022;72(3). doi:10.1099/ijsem.0.005243
40. Verástegui-Valdés MM, Zhang YJ, Rivera-Orduña FN, Cheng HP, Sui XH, Wang ET. Microsymbionts of Phaseolus vulgaris in acid and alkaline soils of Mexico. *Syst Appl Microbiol*. 2014;37(8):605-612. doi:10.1016/j.syapm.2014.08.005
41. Abreu I, Mihelj P, Raimunda D. Transition metal transporters in rhizobia: tuning the inorganic micronutrient requirements to different living styles. *Metallomics*. 2019;11(4):735-755. doi:10.1039/c8mt00372f
42. Shahid M, Singh UB, Khan MS, et al. Bacterial ACC deaminase: Insights into enzymology, biochemistry, genetics, and potential role in amelioration of environmental stress in crop plants. *Front Microbiol*. 2023;14. doi:10.3389/fmicb.2023.1132770
43. Liu KH, Zhang B, Yang BS, et al. Rhizobiales-Specific RirA Represses a Naturally "Synthetic" Foreign Siderophore Gene Cluster To Maintain Sinorhizobium -Legume Mutualism. *mBio*. 2022;13(1). doi:10.1128/mbio.02900-21
44. Cao Y, Achmon Y, Yaron S, Siame BA, Leung KY. Burkholderiaceae and Multidrug Resistance Genes Are Key Players in Resistome Development in a Germfree Soil Model. *mSystems*. 2021;6(6). doi:10.1128/mSystems.00988-21
45. Bromfield ESP, Cloutier S. Complete genome sequences of Rhizobium favelukesii strains T136, T1470, and T1473 from Canada. *Microbiol Resour Announc*. 2026;15(3). doi:10.1128/mra.01239-25
46. González V, Santamaría RI, Bustos P, et al. The partitioned Rhizobium etligenome: Genetic and metabolic redundancy in seven interacting replicons. *Proceedings of the National Academy of Sciences*. 2006;103(10):3834-3839. doi:10.1073/pnas.0508502103
47. Perry BJ, Ferguson S, Laugraud A, Wakelin SA, Reeve W, Ronson CW. Complete Genome Sequences of Trifolium spp. Inoculant Strains Rhizobium leguminosarum sv. trifolii TA1 and CC275e: Resources for Genomic Study of the Rhizobium - Trifolium Symbiosis. *Molecular Plant-Microbe Interactions®*. 2021;34(1):131-134. doi:10.1094/MPMI-08-20-0220-A
48. Jorrin B, Maluk M, Atoliya N, et al. Genomic Diversity of Pigeon Pea (Cajanus cajan L. Millsp.) Endosymbionts in India and Selection of Potential Strains for Use as Agricultural Inoculants. *Front Plant Sci*. 2021;12. doi:10.3389/fpls.2021.680981
49. Chalasani D, Basu A, Pullabhotla SVSRN, et al. Poor Competitiveness of Bradyrhizobium in Pigeon Pea Root Colonization in Indian Soils. *mBio*. 2021;12(4). doi:10.1128/mBio.00423-21

Figures

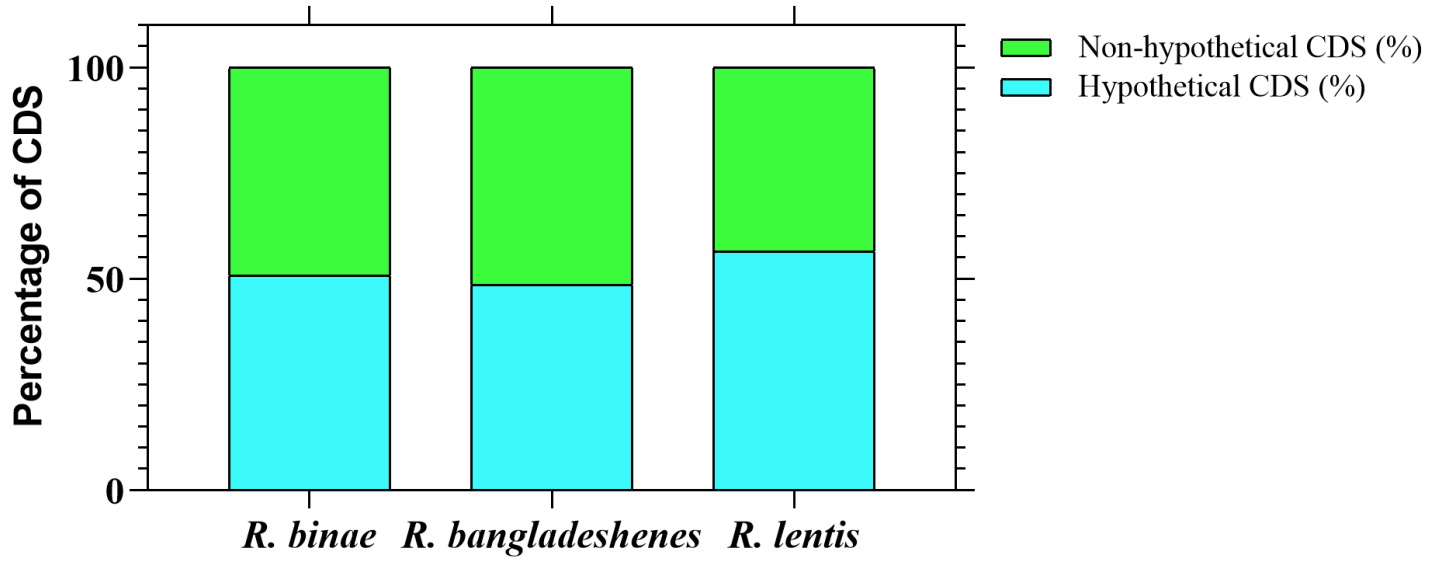


Figure 1

Proportion of hypothetical and non-hypothetical CDS in each genome.

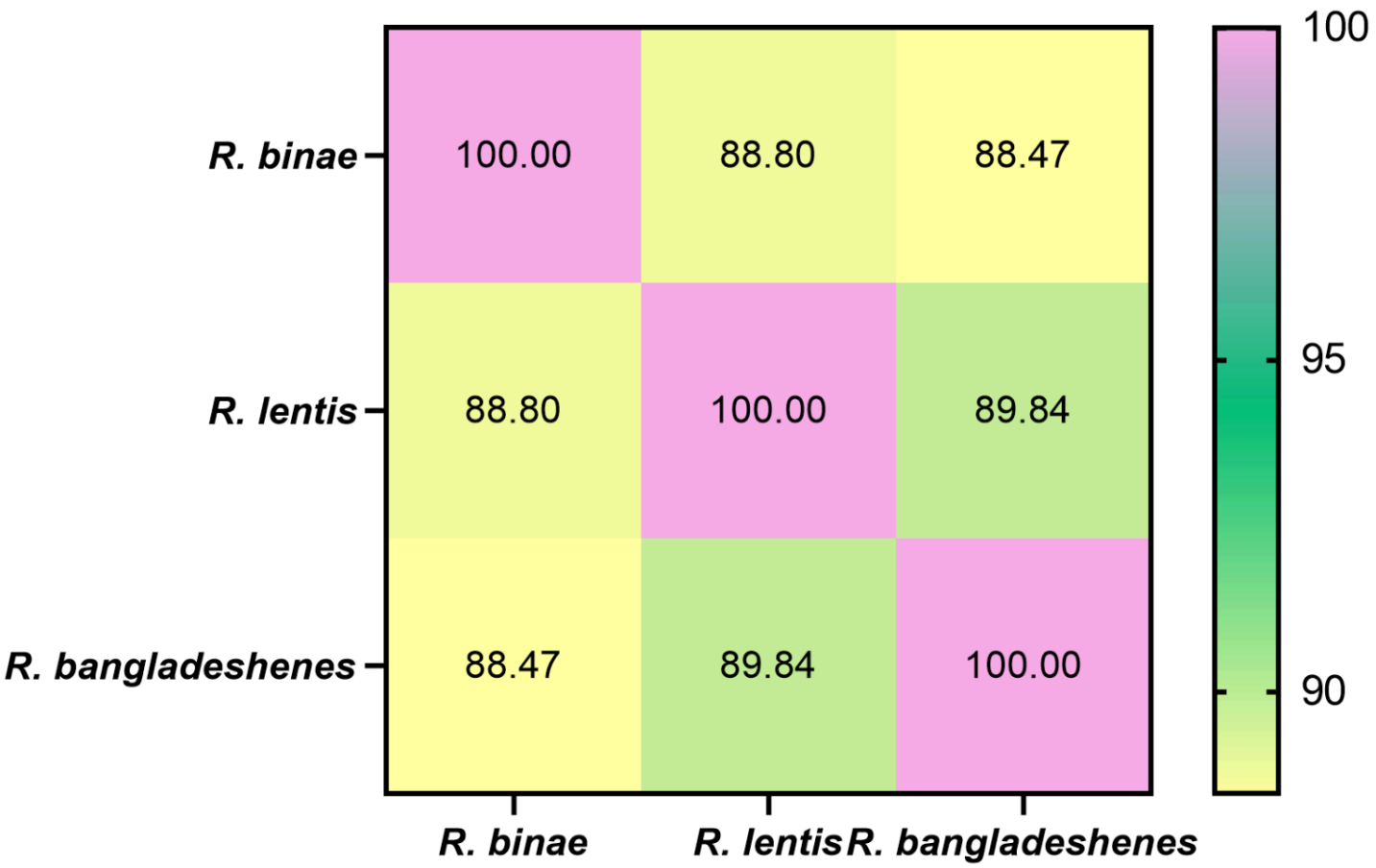


Figure 2

Pairwise average nucleotide identity matrix.

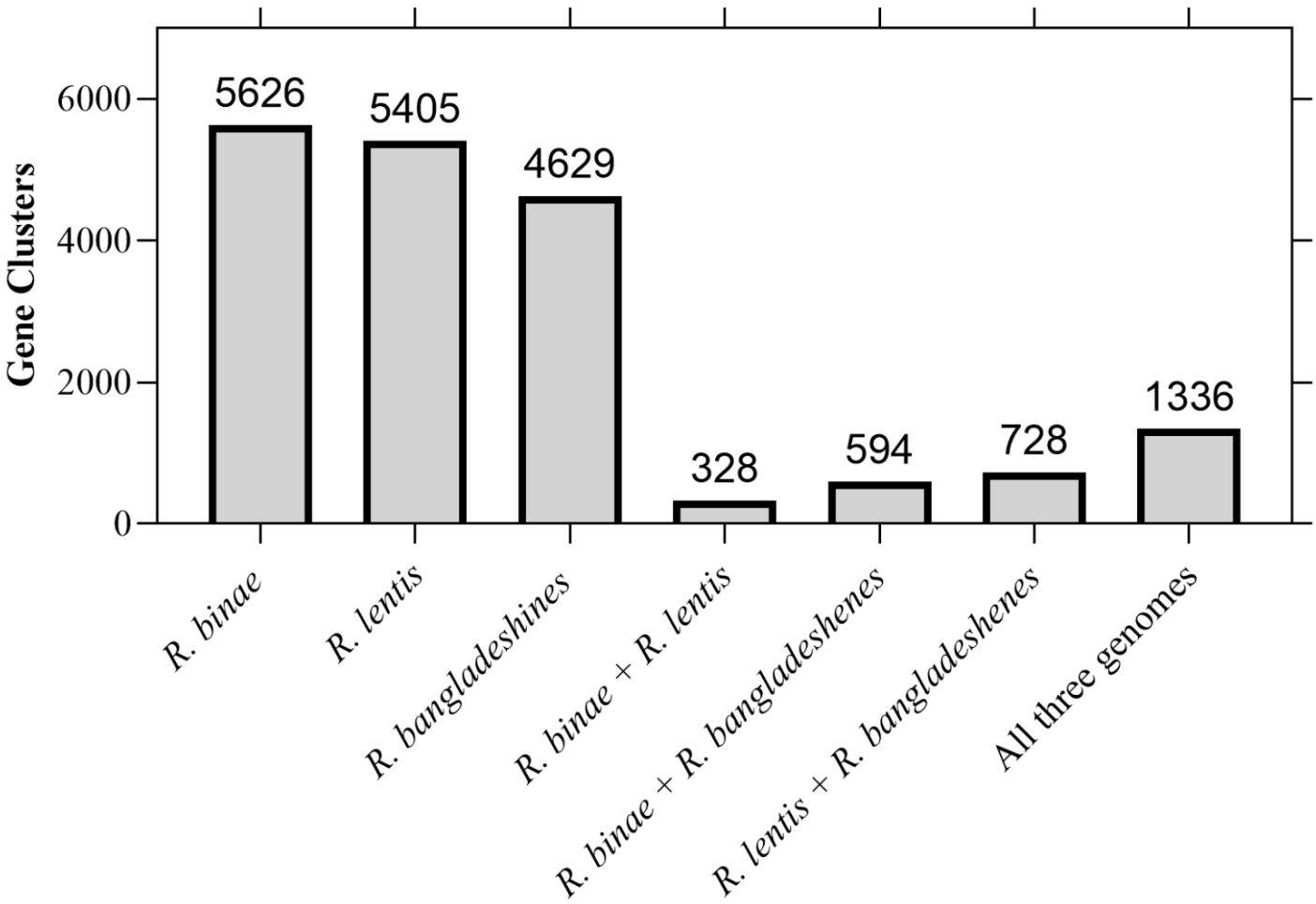


Figure 3

Roary gene-cluster distribution across genome-presence patterns.

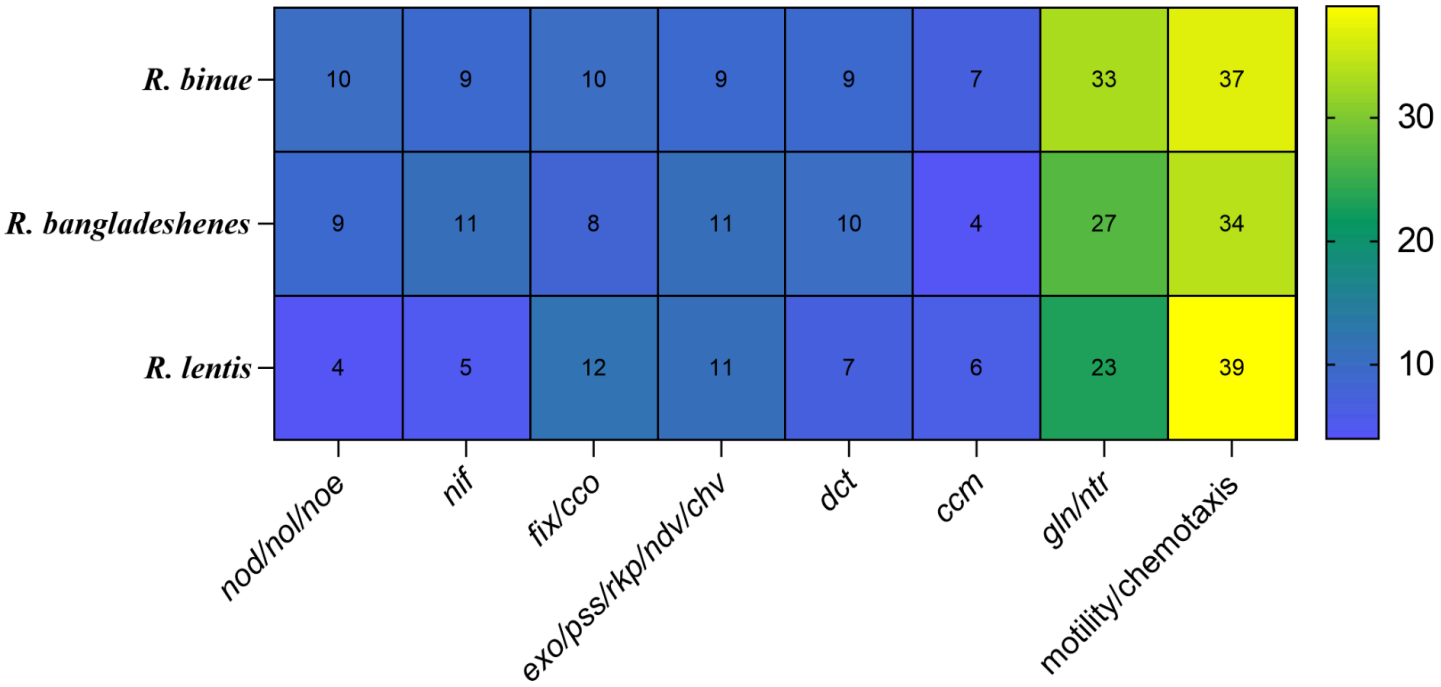


Figure 4

Heatmap of selected symbiosis-associated gene-family counts.

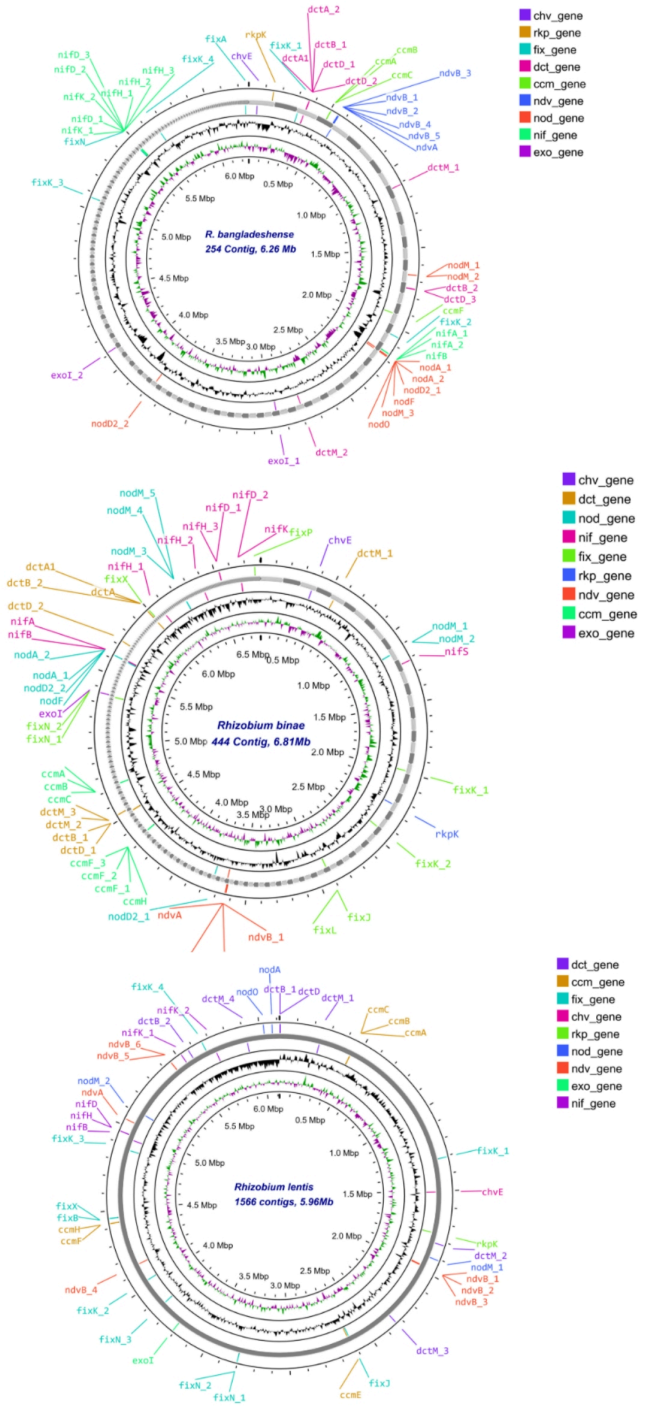


Figure 5

Circular draft genome annotation map of *Rhizobium bangladeshense*, *Rhizobium binae, Rhizobium lentis* showing GC content, GC skew and major symbiosis-associated gene families.

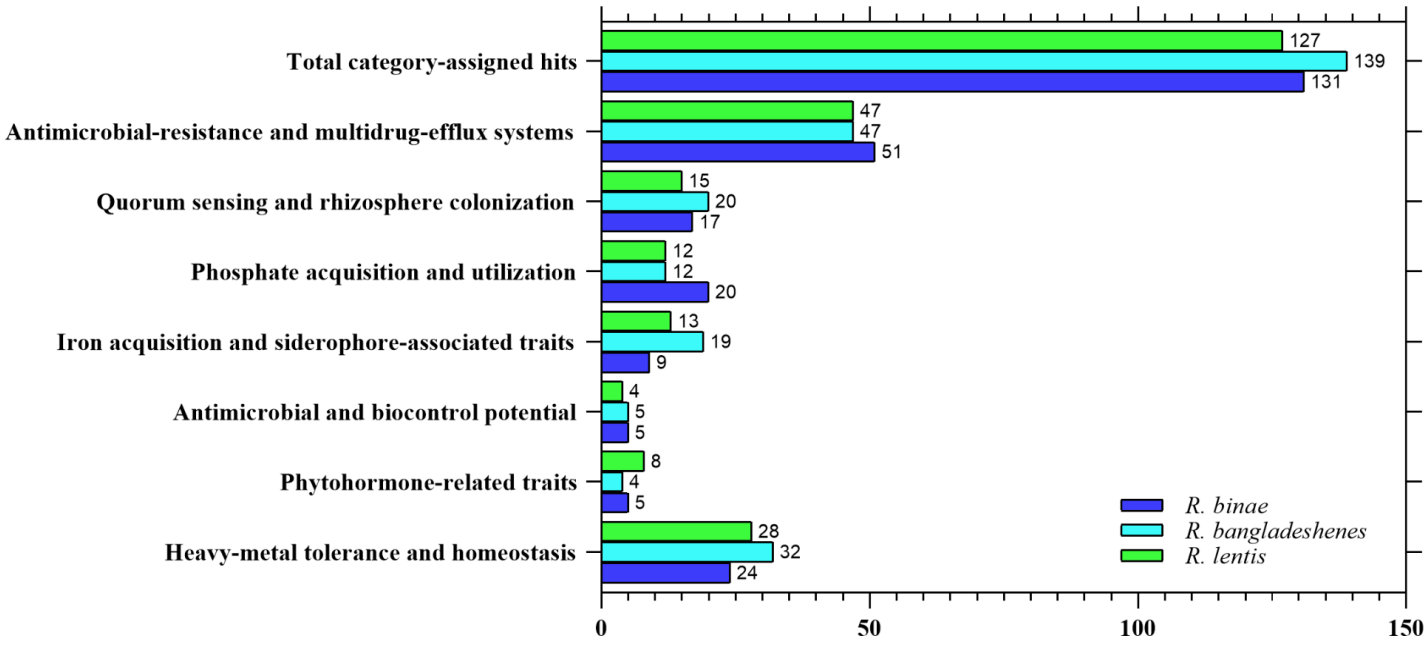


Figure 6

Comparative abundance of beneficial trait and resistance-associated annotation hits.

Comparative genomic analysis of plant growth-promoting and stress-adaptation traits

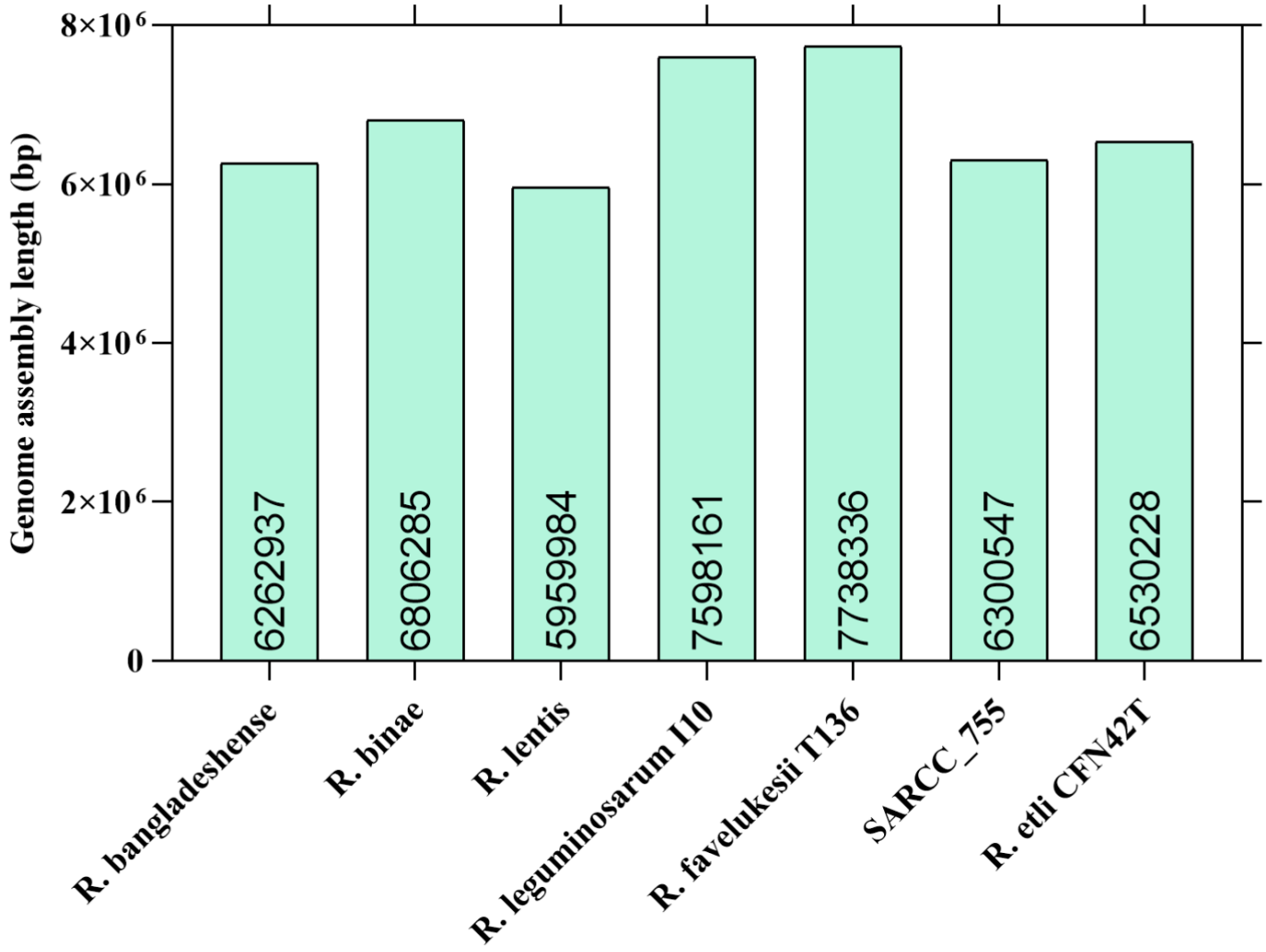


Figure 7

Exact genome assembly lengths in the accession-verified comparison panel.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTableandfigure.docx