Data in Brief 66 (2026) 112764

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Draft genome sequence of *Leuconostoc mesenteroides* isolated from fermented kodo millet: Genomic insights into probiotic and anti-diabetic potential

Navya Sreepathi[a,b], Lakshmi Jayaram[a], Deepthi Puttegowda[a], Chandana Kumari VB[a], Ling Shing Wong[c], Jayanthi MK[b], Ramith Ramu[a,*]

[a] *Department of Biotechnology and Bioinformatics, School of Life Sciences, JSS Academy of Higher Education & Research, Mysuru 570015, Karnataka, India*
[b] *Department of Pharmacology, JSS Medical College, JSS Academy of Higher Education & Research, Mysuru 570015, Karnataka, India*
[c] *Faculty of Health and Life Sciences, INTI International University, Nilai, Malaysia*

ARTICLE INFO

Article history:
Received 16 October 2025
Revised 25 March 2026
Accepted 7 April 2026
Available online 10 April 2026

Keywords:
Leuconostoc mesenteroides
Fermented kodo millet
Whole genome sequencing (WGS)
Probiotic potential
Anti-diabetic potential

ABSTRACT

Leuconostoc mesenteroides RAMULABWGS02, isolated from fermented kodo millet (*Paspalum scrobiculatum*), was subjected to whole-genome sequencing to generate a genomic dataset for functional and comparative analysis. Fermented millet, rich in complex carbohydrates and polyphenols, supports the growth of functional lactic acid bacteria. Genomic DNA was isolated and sequenced on the Illumina NovaSeq 6000 platform, yielding high-quality paired-end reads. The assembled genome spans roughly 2.1 Mb, has a GC content of 37.8%, a sequencing coverage of approximately 1299×, genome completeness of 100.0% (CheckM v1.0.18), and contains plasmid pLEUM1. Genome annotation identified coding sequences, tRNAs, and rRNAs. Functional classification using RAST revealed genes related to carbohydrate metabolism, oxidative stress response, and amino acid biosynthesis. Functional annotation predicted five putative secondary metabolite biosynthetic gene clusters. These genomic pre-

* Corresponding author.
E-mail addresses: ramith.gowda@gmail.com, ramithramu@jssuni.edu.in (R. Ramu).

https://doi.org/10.1016/j.dib.2026.112764

dictions provide a resource for comparative genomics and further functional validation studies The genome sequence is available in GenBank under BioProject PRJNA1254900 and BioSample SAMN48127103.

Specifications Table

Subject	Biology
Specific subject area	Microbial genomics and functional profiling of *Leuconostoc mesenteroides* RAMULABWGS02 with probiotic and predicted anti-diabetic potential.
Type of data	Raw data analyzed
Data collection	The genomic DNA of RAMULABWGS02 was extracted from a pure culture using the phenol-chloroform method. The whole-genome sequencing was performed on the Illumina NovaSeq platform, producing paired-end reads. Quality control assessments and adapter trimming were carried out with FastQC v0.11.9. Phylogenomic analysis was conducted using the Type (Strain) Genome Server (TYGS v408). Trimmed reads were assembled into contigs using SPAdes v3.15.5. Raw reads consisted of 8967,931 paired-end reads generating ~2.69 Gb of sequence data with an estimated genome coverage of ~1299×. A graphical representation of the genome was created using Proksee. Functional annotation of the genome was done using the Rapid Annotation using Systematic Technology (RAST v2.0). The identification of gene clusters responsible for secondary metabolite biosynthesis was achieved through antiSMASH v7.0. Genome completeness and contamination were assessed using CheckM v1.0.18, yielding 100.0% completeness and 1.05% contamination.
Data source location	The strain RAMULABWGS02 was obtained from fermented kodo millet sourced from a local market in Mysore, Karnataka, India.
Data accessibility	The preliminary genome sequence information has been submitted to the NCBI GenBank database, associated with BioProject accession number PRJNA1254900 and BioSample number SAMN48127103. https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA1254900 https://www.ncbi.nlm.nih.gov/biosample/?term=SAMN48127103
Related research article	None

1. Value of the Data

- Provides genomic evidence for the predicted probiotic and anti-diabetic potential of a food-derived LAB strain for further experimental investigation. Whole genome data aids metabolic pathway modeling and discovery of functional gene clusters.
- Useful to researchers exploring probiotics from non-dairy traditional fermented grains.
- Provides a genomic resource as a basis for future experimental studies targeting glucose metabolism and gut health.

2. Background

Type 2 Diabetes Mellitus (T2DM) is a common metabolic disorder characterized by chronically high blood sugar caused by impaired insulin secretion and resistance. With the rising global prevalence, especially in developing countries like India, there is increasing interest in natural, food-based strategies to manage blood glucose levels and related complications [1]. Probiotics, particularly lactic acid bacteria (LAB), have become promising candidates because of their ability to modulate gut microbiota, reduce oxidative stress, and improve glucose metabolism [2]. Traditional fermented foods are valuable sources of these beneficial microbes. Kodo millet (*Paspalum*

scrobiculatum), a fiber-rich, gluten-free ancient grain, provides a unique fermentation matrix rich in bioactive compounds and naturally occurring LAB. During fermentation, microbial populations adapt and produce metabolites that may have potential anti-diabetic effects. Among the LAB, RAMULABWGS02 strain is known for its probiotic functionality and resilience in gastrointestinal conditions [3]. Few studies have explored the genomic features of millet-derived LAB with anti-diabetic properties. In this study, RAMULABWGS02 was isolated from fermented kodo millet and evaluated for its probiotic and enzyme-inhibitory properties. WGS was conducted to uncover the genetic basis of its metabolic traits, especially genes related to carbohydrate metabolism, antioxidant pathways, and the biosynthesis of bioactive compounds potentially relevant to diabetes management.

3. Data Description

The genome draft of RAMULABWGS02 Fig. 1, obtained from fermented kodo millet, and the genome includes a plasmid (pLEUM1) as shown in Fig 2, was sequenced on the Illumina NovaSeq 6000 platform, producing high-quality paired-end reads. The assembled genome measures approximately 2.1 Mb in size and has a GC content of 37.8%. Genome quality metrics assessed by CheckM v1.0.18 confirmed 100% completeness with 1.05% contamination, consistent with a high-quality draft genome. The plasmid pLEUM1 (~37 kbp) harbors several functionally relevant coding sequences. Key resistance genes include cadA (heavy metal efflux), copZ and mco (copper detoxification), arsC (arsenic resistance), crcB_1/crcB_2 (fluoride export), dps (oxidative stress protection), and fetA (iron acquisition), collectively suggesting adaptability in metal-rich environments such as fermented food matrices. Mobile genetic elements include IS3-family transposases (IS1163, IS1520, ISWci2), insK, and the site-specific recombinase pinR. However, no confirmed antibiotic resistance determinants were identified based on current annotation databases. Genome annotation using Rapid Annotation Systematic Technology (RAST) includes a comprehensive array of coding sequences (CDSs), tRNAs, and rRNA shown in Table 1. Additionally, antiSMASH analysis predicted five putative biosynthetic gene clusters, including T1PKS (Type

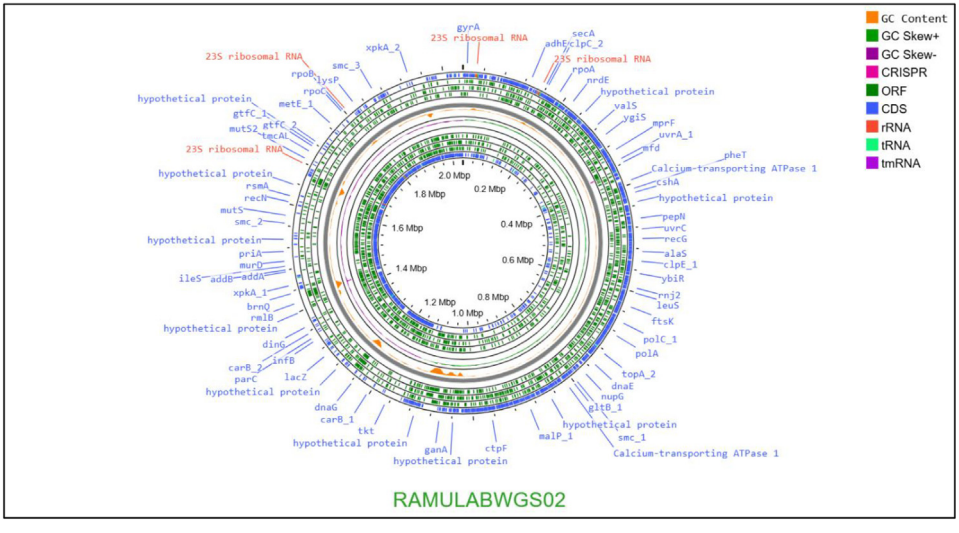

Fig. 1. The circular genome map of RAMULABWGS02 illustrates the features of its draft genome. The outermost ring displays gene names, followed by rings representing coding sequences (CDS) in blue (forward strand) and purple (reverse strand), tRNA in green, rRNA in red, tmRNA in pink, open reading frames (ORFs) as green blocks, CRISPR regions in black, and GC content in orange alongside GC skew in green and purple.

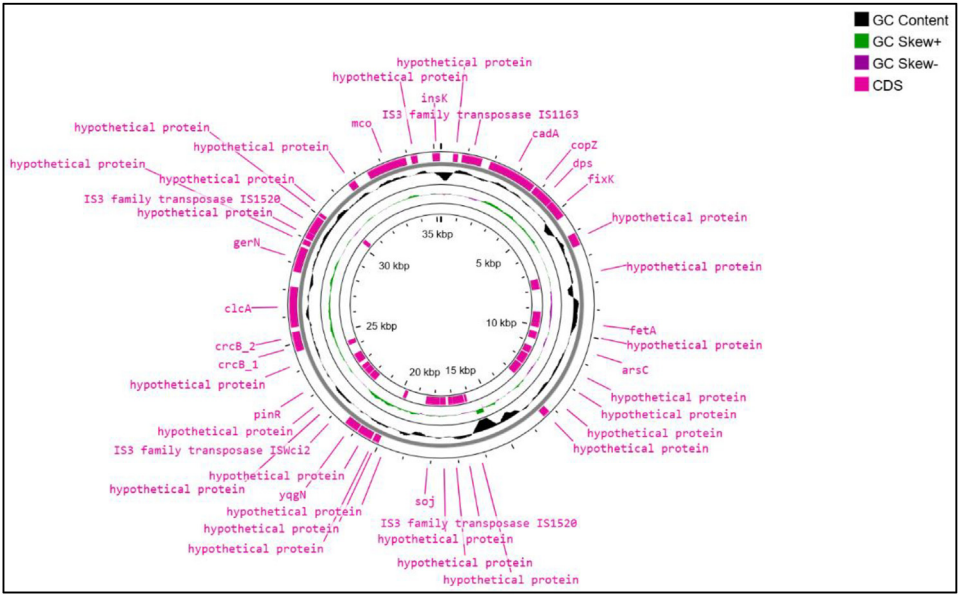

Fig. 2. Circular map of the plasmid pLEUM1 from RAMULABWGS02. The map highlights coding sequences (CDS) depicted in magenta, including key genes such as transposases and metal resistance genes that are labeled for clarity. Inner rings represent the GC content, indicated in black, and the GC skew, illustrated in green and purple, providing insights into the plasmid's genomic characteristics.

Table 1
Genomic profile of *Leuconostoc mesenteroides*.

Feature	Value
Genome	*Leuconostoc mesenteroides*
Domain	Bacteria
Taxonomy ID	1245
Genome size (bp)	2073,653
GC content	37.8
Genome completeness (%)	100.0%
Contamination (%)	1.05%
N50	190,628
L50	3
Number of contigs	2
Plasmid	1
Number of Coding Sequences	1912
Total gene	1995
rRNA	12
tmRNA	1
CRISPR repeat	1

I polyketide synthase), terpene, betalactone, lincosamide-like, and RiPP-like clusters, as shown in Table 2. These predictions indicate potential secondary metabolite biosynthetic capability; however, functional expression and biological activity require further experimental validation. The tree, built using TYGS, confirms clustering with *Leuconostoc mesenteroides* strains shown in Fig 3. Genome-based species confirmation yielded ANIb = 99.25%, ANIm = 99.26% (both above the 95% threshold), and dDDH = 94.3% (above the 70% cutoff), confirming species-level identity. Compared to the type strain L. *mesenteroides* ATCC 8293 (genome size: ~1.97 Mb, GC content: 37.6%), RAMULABWGS02 exhibits a slightly larger genome (2.07 Mb) with a comparable GC con-

Table 2
Biosynthetic gene clusters identified in *Leuconostoc mesenteroides* (RAMULABWGS02).

Cluster type	Base-pair	Biosynthetic categories	Identified secondary metabolites
RiPP-like	66,588 – 78,735	RiPP	Putative lantibiotic or bacteriocin
Lincosamides	470,828 – 536,045	Polyketide / Peptide	Lincosamide-like compound
T3PKS	1159,963 – 1201,120	Polyketide	Flaviolin (putative)
Terpene-precursor	1596,058 – 1616,972	Terpene	Geosmin or Albaflavenone (putative)
Betalactone	1977,471 – 2009,596	Betalactone	-

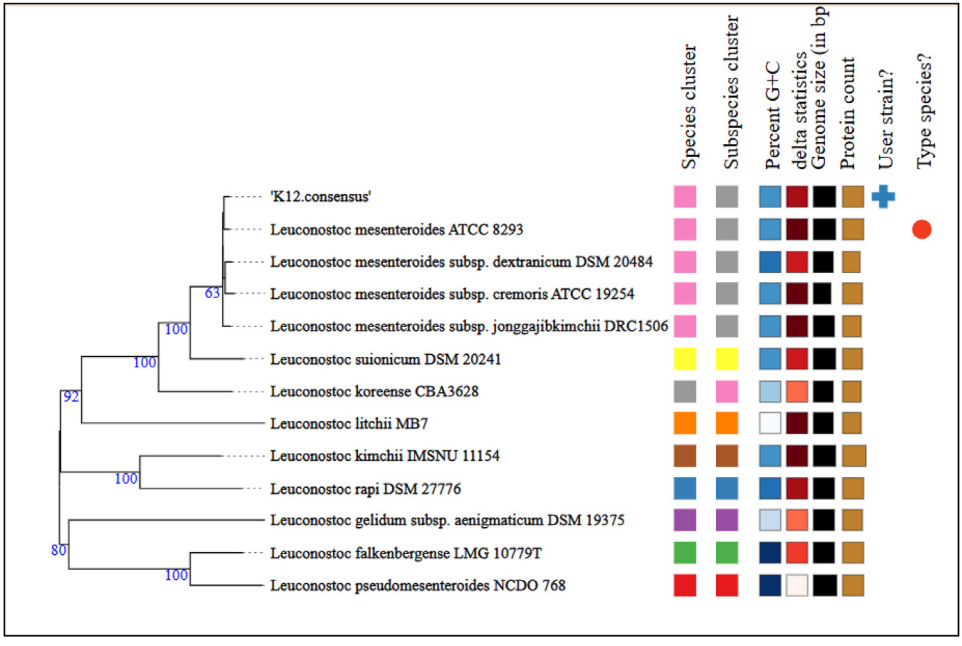

Fig. 3. Phylogenetic tree and genomic features of *Leuconostoc* species. A maximum likelihood tree based on whole-genome data illustrates evolutionary relationships among *Leuconostoc* strains. Bootstrap values (in blue) indicate confidence at branch points. Colored blocks represent comparative genomic features such as species clusters, G + C content, genome size, and protein count. The "+" symbol indicates user-submitted strains, while the red circle marks the type species.

tent of 37.8%, consistent with intraspecies genomic variation reported for this taxon. These genomic features provide a foundational dataset for comparative genomics studies and serve as a basis for future experimental validation of the predicted functional traits of RAMULABWGS02.

4. Experimental Design, Materials and Methods

4.1. Isolation of bacterial strain and preparation of genomic DNA

RAMULABWGS02 was isolated from fermented kodo millet collected in Mysuru, India. Fermentation was carried out by mixing cleaned and shredded kodo millet grains with NaCl (1 tbsp/kg) and incubating the mixture at 37 °C for 48 h with agitation. After fermentation, 1 g of the sample was serially diluted in sterile saline and spread on MRS agar plates. After 48 h of incubation at 37 °C, morphologically distinct colonies were picked and subcultured in MRS broth. Preliminary identification of the isolate was carried out by 16S rRNA gene sequencing [4].

The PCR-amplified 16S rRNA gene was sequenced and subjected to BLAST analysis against the NCBI nucleotide database, providing an initial identification, which was subsequently confirmed and refined through whole-genome sequencing and genome-based taxonomic analysis (ANI and dDDH) in the present study. The selected isolate was preserved in 30% glycerol at −80 °C. For the extraction of genomic DNA, a single colony of RAMULABWGS02 was grown overnight in MRS broth at 37 °C. The cells were collected through centrifugation, and genomic DNA was isolated using a standard phenol-chloroform procedure. The DNA's quality and integrity were verified with a Nanodrop spectrophotometer and 1% agarose gel electrophoresis prior to library preparation and sequencing.

4.2. Genomic assembly and feature annotation

High-quality genomic DNA from RAMULABWGS02 was used to prepare paired-end libraries following the Illumina TruSeq DNA Library Preparation Kit protocol. Whole genome sequencing was performed on the Illumina NovaSeq 6000 system, yielding 150 bp paired-end reads. A total of 8967,931 high-quality read pairs were obtained [5], representing approximately 2.69 Gb of sequence data and achieving a genome coverage of approximately 1299×. The trimmed reads were assembled into contigs using SPAdes v3.15.5 with default parameters and the careful option to reduce mismatches in the final assembly. Annotation was accomplished through the Rapid Annotation Systematic Technology (RAST) [6]. AntiSMASH was utilized to identify gene clusters associated with the biosynthesis of secondary metabolites [7]. A phylogenomic tree was constructed using the Type Strain Genome Server (TYGS) [8]. Genome quality was assessed using CheckM v1.0.18 against the Lactobacillales lineage marker set (350 markers), revealing a completeness of **100.0%** and a contamination estimate of 1.05%

4.3. Genomic similarity analysis using ANI and dDDH

Genome-based taxonomic validation of RAMULABWGS02 was performed using Average Nucleotide Identity (ANI) and digital DNA–DNA hybridization (dDDH) analyses. ANI calculations were conducted using JSpeciesWS employing both BLAST-based (ANIb) and MUMmer-based (ANIm) algorithms. The ANIb value between RAMULABWGS02 and *Leuconostoc mesenteroides* ATCC 8293 (GCA_000014445.1) was **99.25%**, and the ANIm value was **99.26%**, exceeding the established 95% species delineation threshold [9].

Additionally, digital DNA–DNA hybridization (dDDH) values were calculated using the Genome-to-Genome Distance Calculator (GGDC 2.1) via the Type Strain Genome Server (TYGS). The dDDH value (formula d0) between RAMULABWGS02 and the type strain ATCC 8293 was **94.3%**, surpassing the 70% cutoff recommended for species-level identification. The G + C content difference between the genomes was **0.22%**, further supporting taxonomic consistency.

Limitations

The genome reported here is a draft assembly and may contain minor gaps due to short-read sequencing limitations. All functional annotations, including probiotic traits and predicted anti-diabetic properties, are based on in silico tools (RAST, antiSMASH v8.0) and have not been experimentally validated. Biological activities of predicted metabolites require further biochemical confirmation. Strain-specific functional variability among *Leuconostoc mesenteroides* strains means findings may not be broadly generalizable. Genome completeness (100.0%) and contamination (1.05%) were confirmed using CheckM v1.0.18.

Ethics Statement

This study complies with the ethical requirements of *Data in Brief* and does not include human participants, animal testing, or social media data

CRediT Author Statement

Navya Sreepathi: Software, Formal analysis, Investigation; **Lakshmi Jayaram:** Software, Formal analysis, Investigation, Data curation, Visualization, Writing-review & editing, Writing – original draft; **Deepthi Puttegowda:** Software, Formal analysis, Data curation; **Chandana Kumari V B:** Software, Formal analysis, Investigation; **Jayanthi M K:** Super- vision; Writing-review & editing**. Ling Shing Wong:** Supervision, Project administration, Funding acquisition, Writing-review & editing. **Ramith Ramu**: Conceptualization, Project administration, Supervision, Software, Formal analysis, Data curation, Validation, Writing-review & editing.

Data Availability Statement

The draft genome sequence data of *Leuconostoc mesenteroides* are publicly available in the NCBI GenBank repository under BioProject accession number PRJNA1254900 (https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA1254900) and BioSample accession number SAMN48127103 (https://www.ncbi.nlm.nih.gov/biosample/?term=SAMN48127103)

Acknowledgements

All the authors thank JSS Academy of Higher Education & Research (Mysore, Karnataka, India) for their kind support and encouragement.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] S. Divyashree, R. Ramu, M.Y. Sreenivasa, Evaluation of new candidate probiotic lactobacillus strains isolated from a traditional fermented food-multigrain-millet dosa batter, Food Biosci. 57 (2024) 103450, doi:10.1016/j.fbio.2023.103450.

[2] N. Sreepathi, V.B. Chandana Kumari, S.S. Huligere, A. Al-Odayni, V. Lasehinde, M.K. Jayanthi, R. Ramu, Screening for potential novel probiotic levilactobacillus brevis RAMULAB52 with antihyperglycemic property from fermented Carica papaya L, Front. Microbiol. 14 (2023) 1168102, doi:10.3389/fmicb.2023.1168102.

[3] L. Jayaram, R. Ramu, D. Puttegowda, M.K. Jayanthi, V.B. Chandana Kumari, K.W. Goh, A protocol for assessing the adhesion potential of lactic acid bacteria on various cell types, STAR. Protoc. 6 (no. 1) (2025) 103619, doi:10.1016/j.xpro.2025.103619.

[4] B. Gouthami, A. Ramalakshmi, M. Balakrishnan, S. Karthikeyan, I. Muniraj, J.S. Packialakshmi, Functional and molecular characterization of millet associated probiotic bacteria, BMC. Microbiol. 24 (2024) 485 no. 1, doi:10.1186/s12866-024-03606-9.

[5] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, Bioinformatics 30 (no. 15) (2014) 2114–2120, doi:10.1093/bioinformatics/btu170.

[6] R. Overbeek, R. Olson, G.D. Pusch, G.J. Olsen, J.J. Davis, T. Disz, R.A. Edwards, et al., The SEED and the rapid annotation of microbial genomes using Subsystems Technology (RAST), Nucleic. Acids. Res. 42 (2014) D206–D214 no. D1, doi:10.1093/nar/gkt1226.

[7] K. Blin, S. Shaw, H.E. Augustijn, Z.L. Reitz, F. Biermann, M. Alanjary, T. Weber, AntiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation, Nucl. Acid. Res. 51 (2023) W46–W50, doi:10.1093/nar/gkad344.

[8] J.P. Meier-Kolthoff, M. Göker, TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy, Nat. Commun. 10 (2019) 2182, doi:10.1038/s41467-019-10210-3.
[9] M. Richter, R. Rosselló-Móra, F. Oliver Glöckner, J. Peplies, JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison, Bioinformatics 32 (2016) 929–931, doi:10.1093/bioinformatics/btv681.