

Draft genome sequence of *Lactcaseibacillus paracasei* strain LPML01 isolated from the human gut

Hasan Mahfuz Reza,^{1,2} Abu Reza,^{1,2} Md. Ashikuzzaman Antor,^{1,2} Zinia Afrin,^{1,2} Tonima Enam,³ Mahadi Hasan,^{4,5} Afsana Binte Yunus,¹ Md. Shahedur Rahman,^{4,5} Mohammad Abu Hena Mostofa Jamal^{1,2}

AUTHOR AFFILIATIONS See affiliation list on p. 2.

ABSTRACT This study reports the draft genome sequence of *Lactcaseibacillus paracasei* LPML01, a strain of a well-known probiotic species isolated in Bangladesh in 2024. The genome length is 3,061,259 base pairs (bp) with a GC content of 46.13 %.

KEYWORDS Whole genome sequence, *Lactcaseibacillus paracasei*, human gut, probiotics

Lactcaseibacillus paracasei is a well-known probiotic species with potential health-promoting properties (1). This study reports the draft genome sequence of *Lactcaseibacillus paracasei* LPML01, a strain isolated from a human gut sample in 2024, to investigate its probiotic characteristics and genetic features. Because it originates from the human intestinal environment, this strain may be a safe, naturally adapted candidate for probiotic applications.

The stool sample was collected in a sterile collection tube from the 250-bed general hospital, Kushtia. The sample was homogenized in sterile PBS, and serial 10-fold dilutions were plated on MRS agar to isolate lactic acid bacteria, which were incubated for 24 h at 37°C. Although MRS agar is not specifically selective for *L. paracasei*, it is commonly used for the enrichment and preliminary isolation of *Lactobacillus* species (2). A pure culture was prepared from a single colony on a spread culture plate, under the same culture conditions previously mentioned. A single colony from the pure culture was inoculated into MRS broth and incubated overnight at 37°C and 120 rpm. Genomic DNA was extracted from 1 mL broth culture using the CDC PulseNet Total DNA Extraction protocol (https://www.aphl.org/programs/global_health/Documents/PNL33_DNA_Extraction_and_Quality.pdf), which was performed with the Qiagen DNeasy kit to achieve optimal DNA purity. DNA quality was assessed using a Qubit 4.0 Fluorometer and a Nanodrop spectrophotometer, with the Qubit Buffer used at a 1:200 ratio.

Whole-genome sequencing was performed on the Illumina NextSeq 550 platform (paired-end 2 × 150 bp) following the manufacturer's instructions. During library preparation, genomic DNA was enzymatically sheared via tagmentation using the Illumina DNA Prep Reagent Kit, and fragments (200–300 bp) were size-selected with AMPure XP beads. The sequencing library was prepared and adapter-tagged using an automated liquid handler (epMotion 5075). Raw reads were demultiplexed using bcl2fastq v2.20.0 (https://support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software.html).

Raw reads were trimmed with Sickle v1.3 using paired-end trimming set to a length of 20 (3), then assembled with SPAdes v4.0.0 (4). Pilon v1.24 polished the assembly (5). The assembled genome was submitted to PATRIC under BV-BRC v3.54.6 (<https://www.bv-brc.org>) for detailed genomic analysis (6). A whole-genome-based taxonomic analysis was performed on the Type (Strain) Genome Server (<https://tygs.dsmz.de>) (7), with 16S

Editor David Rasko, University of Maryland School of Medicine, Baltimore, Maryland, USA

Address correspondence to Mohammad Abu Hena Mostofa Jamal, jamalbtg@gmail.com, or Md. Shahedur Rahman, ms.rahman@just.edu.bd.

The authors declare no conflict of interest.

See the funding table on p. 3.

Received 22 July 2025

Accepted 20 November 2025

Published 5 December 2025

Copyright © 2025 Reza et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

TABLE 1 Genomic features of *L. paracasei* strain LPML01

Sequencing summary		
Coverage	56.86 ×	
Raw reads	1,306,989	
GC content	45.3%	
Size	163.4 MB	
Assembly report (PATRIC)		
Contigs	37	
GC content	46.13	
Contig L50	3	
Genome length	3,061,259	
Contig N50	339,034	
Genome report (PATRIC)		
Coarse consistency (%)	99.9	
Fine consistency (%)	98.4	
Completeness (%)	100	
Contamination (%)	0	
Annotation report		
Tool used	PGAP	PATRIC
Contigs	37	37
Bases	3,061,259	3,061,259
CDS	2,904	3,109
ncRNA	3	0
rRNA	4	4
Repeat regions	Not detected	21
tRNA	48	48

and whole-genome phylogenetic analyses confirming *L. paracasei* as the species of the isolated strain, designated as LPML01. Default parameters were used for all software unless otherwise specified.

The genome length of *L. paracasei* strain LPML01 is 3,061,259 base pairs, with a GC content of 46.3%. The assembly comprises 37 contigs. The assembled genome was annotated using PATRIC and PGAP v6.10 annotation pipelines (8, 9). Both annotation systems identified 4 rRNA genes; however, PGAP predicted 2,904 coding sequences (CDSs), whereas the PATRIC system identified 3,109 CDS regions. These minor differences likely result from variations in the algorithms, gene prediction criteria, and reference databases used by the two annotation platforms. For instance, PGAP relies on NCBI's curated RefSeq database, whereas PATRIC integrates multiple reference sources (10). A summary of the annotation results is provided in Table 1.

ACKNOWLEDGMENTS

Research grant from the University Grant Commission (Ref No. 37.01.0000.073.03.004.23.257), Fiscal Year: 2021-22, entitled "Molecular Screening of Probiotics from Human Gut Bacteria of Different Aged Child to Evaluate Their Biomedical Application: *In vitro* and Comparative Genomic Analysis."

AUTHOR AFFILIATIONS

¹Department of Biotechnology and Genetic Engineering, Faculty of Biological Sciences, Islamic University, Kushtia, Bangladesh

²Laboratory of Medical and Environmental Biotechnology, Islamic University, Kushtia, Bangladesh

³Department of Pharmacy, Faculty of Biological Sciences, Islamic University, Kushtia, Bangladesh

⁴Department of Genetic Engineering and Biotechnology, Faculty of Biological Sciences and Technology, Jashore University of Science and Technology, Jashore, Bangladesh

⁵Bioinformatics and Microbial Biotechnology Laboratory, Jashore University of Science and Technology, Jashore, Bangladesh

AUTHOR ORCIDS

Abu Reza  <http://orcid.org/0000-0001-5934-4850>

Mahadi Hasan  <https://orcid.org/0009-0000-1821-630X>

Md. Shahedur Rahman  <http://orcid.org/0009-0005-2028-4885>

Mohammad Abu Hena Mostofa Jamal  <http://orcid.org/0000-0003-0542-4323>

FUNDING

Funder	Grant(s)	Author(s)
University Grants Commission of Bangladesh	37.01.0000.073.03.004.23.257	Mohammad Abu Hena Mostofa Jamal

AUTHOR CONTRIBUTIONS

Hasan Mahfuz Reza, Conceptualization, Data curation, Methodology, Validation | Abu Reza, Data curation, Formal analysis, Software, Writing - review and editing | Md. Ashikuzzaman Antor, Data curation, Formal analysis, Methodology, Resources | Zinia Afrin, Formal analysis, Investigation, Resources, Software | Tonima Enam, Methodology, Resources, Software | Mahadi Hasan, Data curation, Formal analysis, Methodology, Software | Afsana Binte Yunus, Formal analysis, Investigation, Methodology | Md. Shahedur Rahman, Investigation, Project administration, Supervision, Validation, Writing – original draft | Mohammad Abu Hena Mostofa Jamal, Conceptualization, Funding acquisition, Project administration, Supervision, Validation, Writing – original draft

DATA AVAILABILITY

This Whole Genome Shotgun project has been deposited in GenBank under accession number [JBNJSA000000000.1](https://www.ncbi.nlm.nih.gov/nuccore/JBNJSA000000000.1) (<https://www.ncbi.nlm.nih.gov/nuccore/JBNJSA000000000.1>). The raw sequence reads are available in the SRA database under accession number [SRR31300363](https://www.ncbi.nlm.nih.gov/sra/SRR31300363) (<https://www.ncbi.nlm.nih.gov/sra/SRR31300363>). Both data are under BioSample number [SAMN48190442](https://www.ncbi.nlm.nih.gov/biosample/SAMN48190442) (<https://www.ncbi.nlm.nih.gov/biosample/SAMN48190442>) and BioProject number [PRJNA1256676](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1256676) (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1256676>). The genome assembly is available under the accession [ASM5011393v1](https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_050113935.1/) (https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_050113935.1/). The assembled genome has been submitted to PATRIC and is available under genome ID [1597.2345](https://www.bv-brc.org/view/Genome/1597.2345) (<https://www.bv-brc.org/view/Genome/1597.2345>).

ETHICS APPROVAL

Ethical approval for this study was obtained from the Dean of the Faculty of Biological Sciences at the Islamic University (Approval Ref: IU/FBS/2024/39), as the university does not have an institutional ethics review committee. Written informed consent was obtained from the minor donor's legal guardian before the collection of the stool sample.

REFERENCES

- Kumaree KK, Prasanth MI, Sivamaruthi BS, Kesika P, Tencomnao T, Chaiyasut C, Prasansuklab A. 2023. *Lactobacillus paracasei* H101 enhances lifespan and promotes neuroprotection in *Caenorhabditis elegans*. *Sci Rep* 13:16707. <https://doi.org/10.1038/s41598-023-43846-9>
- Anonymous. 2003. De man, rogosa and sharpe (MRS) agar, p 511–513. In CorryJEL, CurtisGDW, Baird RM (ed), *Progress in Industrial Microbiology*. Vol. 37. Elsevier.
- Joshi NA FJ. 2011. Sickle: a sliding-window, adaptive, quality-based trimming tool for fastQ files (Version 1.33) [Software]. Available from: <https://github.com/najoshi/sickle>
- Prijbelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. 2020. Using SPAdes *de novo* assembler. *Curr Protoc Bioinformatics* 70:e102. <https://doi.org/10.1002/cpbi.102>
- Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated

- tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. <https://doi.org/10.1371/journal.pone.0112963>
6. Wattam AR, Davis JJ, Assaf R, Boisvert S, Brettin T, Bun C, Conrad N, Dietrich EM, Disz T, Gabbard JL, et al. 2017. Improvements to PATRIC, the all-bacterial bioinformatics database and analysis resource center. *Nucleic Acids Res* 45:D535–D542. <https://doi.org/10.1093/nar/gkw1017>
 7. Meier-Kolthoff JP, Göker M. 2019. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nat Commun* 10:2182. <https://doi.org/10.1038/s41467-019-10210-3>
 8. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res* 44:6614–6624. <https://doi.org/10.1093/nar/gkw569>
 9. Gillespie JJ, Wattam AR, Cammer SA, Gabbard JL, Shukla MP, Dalay O, Driscoll T, Hix D, Mane SP, Mao C, Nordberg EK, Scott M, et al. 2011. PATRIC: the comprehensive bacterial bioinformatics resource with a focus on human pathogenic species. *Infect Immun* 79:4286–4298. <https://doi.org/10.1128/IAI.00207-11>
 10. Haft DH, DiCuccio M, Badretdin A, Brover V, Chetvernin V, O'Neill K, Li W, Chitsaz F, Derbyshire MK, Gonzales NR, Gwadz M, Lu F, et al. 2018. RefSeq: an update on prokaryotic genome annotation and curation. *Nucleic Acids Res* 46:D851–D860. <https://doi.org/10.1093/nar/gkx1068>