

Draft genome sequence of potential probiotic *Lactocaseibacillus paracasei* subsp. *paracasei* S1 isolated from *Nypa fruticans* Wurmb. in Cagayan, Northern Philippines

Lemuel Ray Balloyan,¹ Jimmbeth Zenila Fabia,^{1,2} Ma. Joy Theresa Agcaoili,^{1,2} James Patrick Acang,^{1,3} James Patrick Peralta,^{1,4} Peter James Icalia-Gann^{1,5}

AUTHOR AFFILIATIONS See affiliation list on p. 2.

ABSTRACT Here, we present the draft genome assembly of a putative probiotic *Lactocaseibacillus paracasei* subsp. *paracasei* strain S1 isolated from *Nypa fruticans*. The *L. paracasei* subsp. *paracasei* strain S1 genome is 3.1 Mbp in size, and its annotation revealed the presence of genes involved in several probiotic functionalities.

KEYWORDS whole-genome sequence

Due to numerous health benefits, probiotics have emerged as the most prominent ingredient in the functional food era (1). Here, we used whole-genome sequencing to establish the functionality and safety of our potential probiotic *L. paracasei* subsp. *paracasei* S1 for food and pharmaceutical use. Strain S1 was isolated from *Nypa fruticans* collected from Pamplona, Cagayan (1,827.8470 N°, 12,120.4333 E°). The naturally fermenting sap of the plant was directly streaked onto de Man, Rogosa and Sharpe (MRS) agar. Plates were incubated at 37°C, and colony growth was observed after 24 h. Colonies exhibiting lactic acid bacteria morphology were subjected to subsequent rounds of purification using the same media and growth conditions. Pure cultures of S1 were preserved in 50% glycerol (v/v), stored in –81°C freezer, and are being maintained at the Center for Cellular and Molecular Medical Research in Batac, Ilocos Norte.

The *L. paracasei* subsp. *paracasei* strain S1 was cultured in MRS broth and incubated for 24 h prior to DNA extraction. Approximately 1 mL from the broth culture was aliquoted and centrifuged following the protocols of the GF-1 Bacterial DNA Extraction Kit (Vivantis Technologies, Malaysia). The concentration and quality of the extracted DNA were assessed visually using agarose gel electrophoresis (AGE) and quantified using the spectrophotometry-based ScanDrop² (Analytik Jena, Germany). High-quality DNA was then sent to the DNA Sequencing Core Facility, Philippine Genome Center, Philippines. A DNA library was prepared using the Nextera XT DNA Library Preparation Kit (Illumina Inc., USA) and quality checked using Agilent TapeStation 2200 (Agilent Technologies, USA). Whole-genome sequencing was performed using Illumina MiSeq (Illumina, Inc., USA).

Prior to assembly, the quality of the raw reads was assessed using FastQC v0.12.1 (2). A total of 1,124,122 raw reads were generated with read length ranging from 35 to 151 bp and an expected coverage of 55x. Trimmomatic v0.40 (3) was used to remove adapters and low-quality bases with a quality score cutoff $\geq$ Q20 and minimum length of >36 base pairs. *De novo* assembly of the genome was conducted using SPAdes v4.2.0 (4). Genome completeness and the quality of the assembled genome were assessed using QUAST v5.3.0, BUSCO v6.0.0, and checkM2 v1.1.0 (5–7). The annotation tool Bakta v1.11.3 (8) was then used to annotate the genome with the resulting annotation viewed in the web-based tool Bakta Viewer. Default parameters for the different software were used, except where otherwise noted.

Editor Karthik Anantharaman, University of Wisconsin-Madison, Madison, Wisconsin, USA

Address correspondence to Peter James Icalia-Gann, pjicalia@gmail.com.

The authors declare no conflict of interest.

Received 22 October 2025

Accepted 6 March 2026

Published 16 March 2026

Copyright © 2026 Balloyan 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 Assembly and annotation results of strain S1

Assembly statistics	S1
Number of contigs	41
Genome size (bp)	3,060,919
Largest contig (bp)	651,286
GC content (%)	46.12
N50	350,017
L50	3
Coding density (%)	85%
CheckM2 completeness	99.94%
BUSCO (lactobacillaceae_odb12)	99%
# of proteins	2,867
tRNA	57
rRNA	3

In this study, the draft assembly of strain S1 generated a genome size of 3.1 Mbp with 99% BUSCO score. The genome assembly and annotation results are summarized in Table 1. Submitting the draft genome into the MiGA webserver (9) reveals the top result being *Lactocaseibacillus paracasei* subsp. *paracasei* GCF 004354655.1 (ATCC 25302) having a 98.3% average nucleotide identity (ANI). The presence of genes like *fliK* (10) and *dlt* operon (11) for adhesion, *cps* gene cluster (12) for immunomodulation, *nagB* (13) and *pyrG* for bile tolerance, and BFOOKN_00465 for bacteriocin production was found in the genome. The annotated draft assembly of strain S1 generated in this study will support further studies on its potential for functional food and pharmaceutical use.

ACKNOWLEDGMENTS

The authors would like to acknowledge the financial support of the Philippines' Commission on Higher Education (CHED) through their IFP Probiotics Program. Additionally, the authors would also like to thank the Philippine Genome Center for their technical support in this study.

Lemuel Ray Balloyan, formal analysis, investigation, data curation, and writing—original draft | Jimmbeth Zenila Fabia, investigation and data curation | Ma. Joy Theresa Agcaoili, investigation and data curation | James Patrick Acang, formal analysis | James Patrick Peralta, resources and investigation | Peter James Icalia-Gann, conceptualization, funding acquisition, supervision, and writing—review and editing.

AUTHOR AFFILIATIONS

¹Center for Cellular and Molecular Medical Research, College of Medicine, Mariano Marcos State University, Batac, Ilocos Norte, Philippines

²Research Directorate, Mariano Marcos State University, Batac, Ilocos Norte, Philippines

³Computer Science Department, Mariano Marcos State University, Batac, Ilocos Norte, Philippines

⁴Department of Medical Microbiology, College of Medicine, Mariano Marcos State University, Batac, Ilocos Norte, Philippines

⁵Department of Biochemistry and Molecular Biology, College of Medicine, Mariano Marcos State University, Batac, Ilocos Norte, Philippines

AUTHOR ORCIDs

Peter James Icalia-Gann  <http://orcid.org/0009-0003-5342-4464>

AUTHOR CONTRIBUTIONS

Lemuel Ray Balloyan, Data curation, Formal analysis, Investigation, Methodology, Validation, Writing – original draft, Writing – review and editing | Jimmbeth Zenila Fabia,

Conceptualization, Data curation, Project administration, Resources | Ma. Joy Theresa Agcaoli, Conceptualization, Project administration, Resources | James Patrick Acang, Formal analysis, Investigation, Validation | James Patrick Peralta, Conceptualization, Investigation, Project administration, Resources | Peter James Icalia-Gann, Conceptualization, Data curation, Funding acquisition, Investigation, Project administration, Resources, Supervision, Writing – review and editing

DATA AVAILABILITY

The whole genome data of *Lactocaseibacillus paracasei* subsp. *paracasei* S1 can be obtained under accession [JBTNPG000000000.1](https://www.ncbi.nlm.nih.gov/nuclseq/JBTNPG000000000.1) in GenBank. The BioSample and BioProject accession numbers are [SAMN50018499](https://www.ncbi.nlm.nih.gov/biosample/SAMN50018499) and [PRJNA1293312](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1293312), respectively. The raw sequences can be accessed in NCBI with SRA accession [SRR35844808](https://www.ncbi.nlm.nih.gov/sra/SRR35844808).

REFERENCES

1. Latif A, Shehzad A, Niazi S, Zahid A, Ashraf W, Iqbal MW, Rehman A, Riaz T, Aadil RM, Khan IM, Özogul F, Rocha JM, Esatbeyoglu T, Korma SA. 2023. Probiotics: mechanism of action, health benefits and their application in food industries. *Front Microbiol* 14:1216674. <https://doi.org/10.3389/fmicb.2023.1216674>
2. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. Available from: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
3. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
4. Prjibelski 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>
5. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29:1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>
6. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>
7. Chklovski A, Parks DH, Woodcroft BJ, Tyson GW. 2023. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. *Nat Methods* 20:1203–1212. <https://doi.org/10.1038/s41592-023-01940-w>
8. Schwengers O, Jelonek L, Dieckmann MA, Beyvers S, Blom J, Goesmann A. 2021. Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. *Microb Genom* 7:11. <https://doi.org/10.1099/mgen.0.000685>
9. Rodríguez-R LM, Gunturu S, Harvey WT, Rosselló-Mora R, Tiedje JM, Cole JR, Konstantinidis KT. 2018. The microbial genomes atlas (MiGA) webserver: taxonomic and gene diversity analysis of archaea and bacteria at the whole genome level. *Nucleic Acids Res* 46:W282–W288. <https://doi.org/10.1093/nar/gky467>
10. Haiko J, Westerlund-Wikström B. 2013. The role of the bacterial flagellum in adhesion and virulence. *Biology (Basel)* 2:1242–1267. <https://doi.org/10.3390/biology2041242>
11. Weidenmaier C, Peschel A. 2008. Teichoic acids and related cell-wall glycopolymers in gram-positive physiology and host interactions. *Nat Rev Microbiol* 6:276–287. <https://doi.org/10.1038/nrmicro1861>
12. Li J, Feng S, Yu L, Zhao J, Tian F, Chen W, Zhai Q. 2022. Capsular polysaccharides of probiotics and their immunomodulatory roles. *Food Science and Human Wellness* 11:1111–1120. <https://doi.org/10.1016/j.fshw.2022.04.003>
13. Liu C, Xu L, Chen S, Wang S, Lv M, Dong K, Lin Q, Cao Z. 2025. Whole-genome sequencing and probiotic properties of *Lactobacillus helveticus* KM7 isolated from the gut of the Chinese honey bee (*Apis cerana*): a promising exopolysaccharide-producing strain. *BMC Microbiol* 25:530. <https://doi.org/10.1186/s12866-025-04286-9>