



# Draft Genome Sequences of Seven *Limosilactobacillus fermentum* Indigenously Isolated Probiotic Strains from the Artisanal Fermented Milk Product Dahi

Muhammad Nadeem Khan,<sup>a</sup>  Muhammad Imran<sup>a</sup>

<sup>a</sup>Department of Microbiology, Quaid-i-Azam University, Islamabad, Pakistan

**ABSTRACT** Here, we report the draft genome sequences of seven strains of potentially probiotic *Limosilactobacillus fermentum* isolated from the traditional fermented milk product dahi. The estimated average genome size was 1,955,815 bp, with a median GC content of 52%. Genome annotation predicted an average of 1,871 protein-coding genes and 47 RNAs.

Lactic acid bacteria (LAB) are known for their ability to impart a multitude of health benefits. Recently, *Lactobacillus fermentum*, a member of LAB, was reclassified as *Limosilactobacillus fermentum* for its metabolic and ecological properties (1). *L. fermentum* is a Gram-positive, non-spore-forming, nonmotile, rod-shaped bacterium that can grow singly, in pairs, or in short chains (1). *L. fermentum* is usually found in habitats that are nutrient rich and are associated with humans, plants, animals, and food (2). *L. fermentum* is considered a promising potential probiotic candidate due to its lack of antibiotic resistance genes, along with its antioxidative, antimicrobial, and cholesterol reduction properties (3–5). The genome of *L. fermentum* possesses specific genes that help it to tolerate the immune system of the host, interact with the other microbes of the gut, and colonize the host epithelium (6–8).

We report the genome sequences of seven strains of *L. fermentum* (QAULFN56, QAULFN64, QAULFN21, QAULFN53, QAULFN54, QAULFN55, and QAULFN62) that were isolated from the traditional fermented milk product dahi. Samples were collected from dairy corner shops of Rawalpindi, Islamabad, Pakistan. For isolation, 100  $\mu$ l of dahi sample was plated on de Man Rogosa Sharpe (MRS) agar (pH 6.8  $\pm$  2) and incubated at 37°C for 24 h in a GasPak anaerobic system (Sigma-Aldrich). Single colonies from MRS plates were subcultured, and purified cultures were obtained. The cultures were identified as strains of *L. fermentum* with a microbial identification system explained previously (9), followed by analysis using the ON-rep-seq gene sequencing tool (10).

The pure colonies from agar plates were grown overnight in MRS broth followed by DNA extraction using InstaGene matrix (Bio-Rad Laboratories, California, USA). A Qubit 2.0 fluorometer (Invitrogen, Carlsbad, CA, USA) was used to quantify the DNA. DNA libraries for high-throughput sequencing were obtained with the commercially available Vazyme TruePrep DNA library prep kit V2 (Vazyme Biotech, Nanjing, China). Sequencing of these libraries was performed on an Illumina HiSeq-2000 sequencing platform (BGI-Shenzhen, Shenzhen, China) with an average read length of 101 bp. FASTQ files were generated, and average total raw reads were 10,294,243. Default parameters were used for all software unless otherwise specified. The quality trimming of the sequencing reads was performed using Trimmomatic version 0.38 (11). The number of filtered average reads was 9,340,781, and these were aligned with the reference genome of *L. fermentum* IFO 3956 (NCBI RefSeq accession number [NC\\_010610](https://.ncbi.nlm.nih.gov/nuccore/NC_010610)). The assembly of the high-quality reads was performed using the Velvet (v1.2.10) assembler (12). The *de novo* assembly results were evaluated using Quast software (13). The assembled

**Editor** Julie C. Dunning Hotopp, University of Maryland School of Medicine

**Copyright** © 2022 Khan and Imran. 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/).

Address correspondence to Muhammad Imran, [mmimran@qau.edu.pk](mailto:mmimran@qau.edu.pk).

The authors declare no conflict of interest.

**Received** 20 July 2022

**Accepted** 21 September 2022

**Published** 10 October 2022

**TABLE 1** Genome statistics for *L. fermentum* isolates obtained from the fermented milk product dahi

Feature	QAULFN56	QAULFN64	QAULFN21	QAULFN53	QAULFN54	QAULFN55	QAULFN62
GenBank accession no.	JAJAOT000000000	JAJAOV000000000	JAJAOP000000000	JAJAQ000000000	JAJAOR000000000	JAJAOS000000000	JAJAOU000000000
SRA accession no.	SRR19959948	SRR19959946	SRR19959952	SRR19959951	SRR19959950	SRR19959949	SRR19959947
Raw reads	10,296,951	10,294,952	10,293,176	10,292,199	10,297,350	10,292,921	10,292,151
Genome coverage	300×	300×	300×	300×	300×	300×	300×
Size (bp)	1,900,053	1,900,053	1,867,005	2,077,539	1,966,551	2,011,311	1,968,193
% GC	52.4	52.4	52.8	51.7	52	51.8	52
Contigs	102	102	74	170	129	137	80
Total genes	1,919	1,919	1,848	2,148	2,059	2,041	2,001
CDSs with proteins	1,803	1,803	1,724	2,002	1,965	1,935	1,869
RNAs	52	52	63	53	20	29	60
Pseudogenes	64	64	61	80	74	70	72
Subsystem	207	207	212	220	215	220	213
COGs	776	778	795	845	826	840	822
eggNOG features	1,653	1,643	1,595	1,800	1,715	1,749	1,684
Bacteriocin gene	Enterolysin_A	Enterolysin_A	Enterolysin_A	Enterolysin_A	Enterolysin_A	Enterolysin_A	Enterolysin_A
CRISPR <sup>a</sup>	C	C	C	P	Nil	C	C

<sup>a</sup> P, possible; C, confirmed. No resistance gene was found in any of the strains.

genomes were annotated through PGAP (14). The subsystem identification in the genomes was determined through RAST (15), while evaluation of orthologous genes was performed via clusters of orthologous genes (COGs) analysis (16) and eggNOG (17). The results are presented in Table 1. The genomes were mined for the presence of bacteriocin, antibiotic resistance, and CRISPR/CRISPR-Cas genes through the online tools BAGEL4, CARD, and CRISPR Finder, respectively (18–22). Except for strains QAULFN54 and QAULFN21, the rest of the five strains had bacteriocin genes. CRISPR sequences were identified in all the strains except QAULFN54. The presence of CRISPRs suggested immunity to foreign attacking agents, like phages, jumping genes, or transfer elements (23).

**Data availability.** The genomic data reported here have been deposited in GenBank under the accession numbers given in Table 1. The SRA accession numbers are also provided in Table 1. The BioProject number is [PRJNA744373](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA744373), and the relevant BioSample numbers are [SAMN20114169](https://www.ncbi.nlm.nih.gov/biosample/SAMN20114169) to [SAMN20114175](https://www.ncbi.nlm.nih.gov/biosample/SAMN20114175).

## ACKNOWLEDGMENT

This work was supported by Higher Education Commission, Pakistan (HEC) under NRPJ Project no 6167.

## REFERENCES

- Zheng J, Wittouck S, Salvetti E, Franz CMAP, Harris HMB, Mattarelli P, O'Toole PW, Pot B, Vandamme P, Walter J, Watanabe K, Wuyts S, Felis GE, Gänzle MG, Lebeer S. 2020. A taxonomic note on the genus *Lactobacillus*: description of 23 novel genera, emended description of the genus *Lactobacillus* Beijerinck 1901, and union of *Lactobacillaceae* and *Leuconostocaceae*. *Int J Syst Evol Microbiol* 70:2782–2858. <https://doi.org/10.1099/ijsem.0.004107>.
- Duar RM, Lin XB, Zheng J, Martino ME, Grenier T, Pérez-Muñoz ME, Leulier F, Gänzle M, Walter J. 2017. Lifestyles in transition: evolution and natural history of the genus *Lactobacillus*. *FEMS Microbiol Rev* 41:527–548. <https://doi.org/10.1093/femsre/fux030>.
- Mikelsaar M, Sepp E, Stsepetova J, Hutt P, Zilmer K, Kullisaar T, Zilmer M. 2015. Regulation of plasma lipid profile by *Lactobacillus fermentum* (probiotic strain ME-3 DSM14241) in a randomised controlled trial of clinically healthy adults. *BMC Nutr* 1:27. <https://doi.org/10.1186/s40795-015-0020-z>.
- Naghmouchi K, Belguesmia Y, Bendali F, Spano G, Seal BS, Drider D. 2020. *Lactobacillus fermentum*: a bacterial species with potential for food preservation and biomedical applications. *Crit Rev Food Sci Nutr* 60:3387–3399. <https://doi.org/10.1080/10408398.2019.1688250>.
- Ghosh T, Beniwal A, Semwal A, Navani NK. 2019. Mechanistic insights into probiotic properties of lactic acid bacteria associated with ethnic fermented dairy products. *Front Microbiol* 10:502. <https://doi.org/10.3389/fmicb.2019.00502>.
- Garcia-Gonzalez N, Battista N, Prete R, Corsetti A. 2021. Health-promoting role of *Lactiplantibacillus plantarum* isolated from fermented foods. *Microorganisms* 9:349. <https://doi.org/10.3390/microorganisms9020349>.
- Fidanza M, Panigrahi P, Kollmann TR. 2021. *Lactiplantibacillus plantarum*: nomad and ideal probiotic. *Front Microbiol* 12:712236. <https://doi.org/10.3389/fmicb.2021.712236>.
- Filannino P, De Angelis M, Di Cagno R, Gozzi G, Riciputi Y, Gobbetti M. 2018. How *Lactobacillus plantarum* shapes its transcriptome in response to contrasting habitats. *Environ Microbiol* 20:3700–3716. <https://doi.org/10.1111/1462-2920.14372>.
- Klingler JM, Stowe RP, Obenhuber DC, Groves TO, Mishra SK, Pierson DL. 1992. Evaluation of the Biolog automated microbial identification system. *Appl Environ Microbiol* 58:2089–2092. <https://doi.org/10.1128/aem.58.6.2089-2092.1992>.
- Krych Ł, Castro-Mejia JL, Forero-Junco LM, Moesby DN, Mikkelsen MB, Rasmussen MA, Sykulski M, Nielsen DS. 2019. DNA enrichment and tagmentation method for species-level identification and strain-level differentiation using ON-rep-seq. *Commun Biol* 2:369. <https://doi.org/10.1038/s42003-019-0617-x>.
- 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>.
- Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18:821–829. <https://doi.org/10.1101/gr.074492.107>.
- 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>.
- 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>.
- Aziz RK, Bartels D, Best AA, Dejongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL,

- Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 9:75. <https://doi.org/10.1186/1471-2164-9-75>.
16. Galperin MY, Makarova KS, Wolf YI, Koonin EV. 2015. Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res* 43:D261–D269. <https://doi.org/10.1093/nar/gku1223>.
  17. Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, Rattei T, Mende DR, Sunagawa S, Kuhn M, Jensen LJ, von Mering C, Bork P. 2016. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res* 44:D286–D293. <https://doi.org/10.1093/nar/gkv1248>.
  18. van Heel AJ, de Jong A, Song C, Viel JH, Kok J, Kuipers OP. 2018. BAGEL4: a user-friendly web server to thoroughly mine RiPPs and bacteriocins. *Nucleic Acids Res* 46:W278–W281. <https://doi.org/10.1093/nar/gky383>.
  19. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, Huynh W, Nguyen A-LV, Cheng AA, Liu S, Min SY, Miroshnichenko A, Tran H-K, Werfalli RE, Nasir JA, Oloni M, Speicher DJ, Florescu A, Singh B, Faltyn M, Hernandez-Koutoucheva A, Sharma AN, Bordeleau E, Pawlowski AC, Zubyk HL, Dooley D, Griffiths E, Maguire F, Winsor GL, Beiko RG, Brinkman FSL, Hsiao WWL, Domselaar GV, McArthur AG. 2020. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res* 48:D517–D525.
  20. Grissa I, Vergnaud G, Pourcel C. 2008. CRISPRcompar: a website to compare clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res* 36:W145–W148. <https://doi.org/10.1093/nar/gkn228>.
  21. Grissa I, Vergnaud G, Pourcel C. 2007. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* 8:172. <https://doi.org/10.1186/1471-2105-8-172>.
  22. Grissa I, Vergnaud G, Pourcel C. 2007. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res* 35:W52–W57. <https://doi.org/10.1093/nar/gkm360>.
  23. Marraffini L. 2015. CRISPR-Cas immunity in prokaryotes. *Nature* 526:55–61. <https://doi.org/10.1038/nature15386>.