Data Article

Draft genome sequence data of *Lactiplantibacillus plantarum* 33C isolated from Lithuanian fermented food

Ashwinipriyadarshini Megur [a],*, Eglė Lastauskienė [b], Aurelijus Burokas [a],*

[a] *Department of Biological Models, Institute of Biochemistry, Life Sciences Center, Vilnius University, Sauletekio Ave. 7, LT, Vilnius 10257, Lithuania*
[b] *Department of Microbiology and Biotechnology, Institute of Biosciences, Life Sciences Center, Vilnius University, Sauletekio Ave. 7, LT, Vilnius 10257, Lithuania*

ARTICLE INFO

ABSTRACT

This strain was isolated from traditionally (homemade) fermented Lithuanian cherry tomatoes. The genome consists of 55 contigs with a total size of 3,326,119 bp, an N50 of 170738, and a GC% of 44.3 %. According to the COG annotation, most of these proteins were divided into three categories related to transcription (K category: 307), amino acid transport and metabolism (E category: 222), and carbohydrate transport and metabolism (G category: 268). No genes associated with antimicrobial resistance or virulence factors were identified. The data presented here can be used in comparative genomics to identify antimicrobial resistance genes and virulence factors that may be present in relevant *Lactobacillus* species. PlasmidFinder server revealed the presence of plasmid pR18 (assessment number JN601038) in the genome that has lincomycin resistance, can be transferred from one bacterium to another, and is one of the most common plasmids in the genera *Bacillus* and *Lactobacillus*. The draft genome sequence data have been deposited with NCBI under Bioproject under accession number PRJNA947394.

© 2024 The Author(s). Published by Elsevier Inc.
This is an open access article under the CC BY-NC license
(http://creativecommons.org/licenses/by-nc/4.0/)

* Corresponding authors.
E-mail addresses: ashwinipriyadarshini.megur@bchi.stud.vu.lt (A. Megur), aurelijus.burokas@gmc.vu.lt (A. Burokas).
Social media: @ashwini_megur (A. Megur)

Specifications Table

| Subject	Microbiology Genetics
Specific subject area	Microbial genomics and beneficial bacteria
Type of data	Raw reads of sequenced genome, assembled and annotated draft genome
Data collection	Pure culture of *Lactoplantibacillus plantarum* 33C was used to isolate genomic DNA according to the Qiagen Genomic DNA Handbook (Qiagen). The NovaSeq 6000 Platform (Illumina) was used to sequence the genomic DNA. The raw reads were used for genome assembly, and the annotation, search of antimicrobial resistance genes, virulence factors, and plasmids were based on the assembled genome.
Data source location	Institution: Department of Biological models, Institute of Biochemistry, Life Science Center, Vilnius University City/Town/Region: Vilnius Country: Lithuania
Data accessibility	Repository name: National Center for Biotechnology Information (NCBI) Data identification number: BioProject: PRJNA947394, BioSample: SAMN33849716, GenBank: JARUWW000000000 Direct URL to data: https://www.ncbi.nlm.nih.gov/nuccore/JARUWW000000000 https://www.ncbi.nlm.nih.gov/bioproject/PRJNA947394 https://www.ncbi.nlm.nih.gov/biosample/SAMN33849716
Related research article	A. Megur, E.B.M. Daliri, T. Balnionytė, J. Stankevičiūtė, E. Lastauskienė, A. Burokas, In vitro screening, and characterization of lactic acid bacteria from Lithuanian fermented food with potential probiotic properties, Front Microbiol 14 (2023). https://doi.org/10.3389/fmicb.2023.1213370.

1. Value of the Data

- *Lactiplantibacillus plantarum* is a lactic acid bacteria commonly used for food fermentation and to improve human and animal health by gut modulation. The sequencing data described here unveils the identity and probiotic characteristics of *L. plantarum* 33C, a valuable fermented food isolate.
- The information provided in this dataset could be beneficial for researchers studying genomic analysis of lactic acid bacteria and food fermentations.
- The information provided for *L. plantarum* 33C is noteworthy because this bacterium does not possess any of the antimicrobial resistance genes like macB, vanHD, vanL, baeR, baeS, mfd, lmrC, lmrD, salA, lsaA, rpsJ, tetT, tetM, efrA, efrB, tetA, vanL, TaeA, vgaD and rpsJ commonly found in other *L. plantarum* isolates studied in the past [1,2].
- Whole genome sequencing data can be utilized to evaluate the safety of using L. plantarum 33C as a probiotic or starter culture in the functional dairy food industry.

2. Background

L. plantarum 33C was isolated from Lithuanian fermented cherry tomatoes procured from Kalvariju market, Vilnius, Lithuania [3]. The strain was isolated from pure morphological selection on Man Ragosa Sharpe agar and sent for whole genome sequencing. In a series of in-vitro experiments, we observed that the strain has the probiotic abilities to survive in gastrointestinal conditions, adhesion on HCT116 human colon cells and ability to produce antimicrobial properties against pathogenic bacteria [3]. Here, we have reported the genome sequence of strain 33C to allow its identification at the genus and species level and facilitate further molecular studies for its safety to obtain insights into possible antimicrobial resistance genes and virulent factors it might possess.

3. DATA Description

In this study, we present the complete genome sequencing data for *L. plantarum* 33C, as well as an analysis of its genome data, taxonomic identification, antimicrobial resistance, and the presence of virulence factors for safety assessment. The genome consisted of 55 contigs
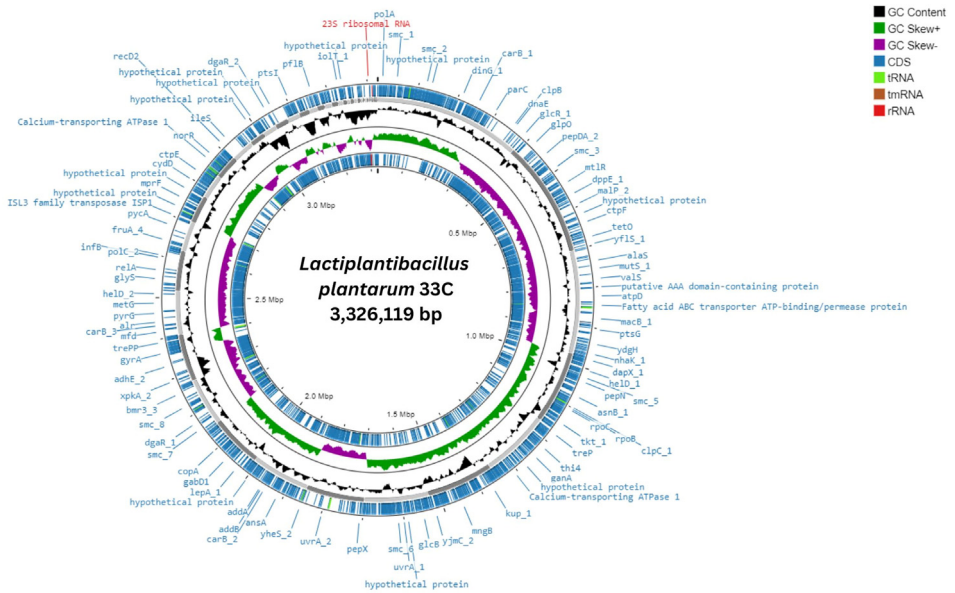
Fig. 1. Genome map of *L. plantarum* 33C built using CGView server (https://proksee.ca/ assessed on 5th April 2024). Genomic features are represented from the outer to the inner circle, Circle 1- illustrates Prokka annotated forward and Coding sequences (CDS) with tRNA, rRNA, and tmRNA; Circle 2- represents the GC content; Circle 3- displays the GC skew (G-C) / (G+C); and Circle 4 shows the genome size- 3326,119bp.

with a total size of 3326,119 bp, an N50 of 170738 and a GC% of 44.3 %. The Bakta annotation produced 3208 genes, of which 3145 were CDS (coding sequences), 3 rRNAs (ribosomal RNAs), 1 tmRNA (transfer-messenger RNA) and 61 tRNAs (transfer RNA) (Fig. 1). s. The final set of core-genome SNPs was used to rebuild the phylogenomic relationship among the neighbouring genome using FastTree2, revelling the close relation to *L. plantarum* species (Fig. 2). Further analysis revealed Cluster of Orthologous Groups (COG) functional category. Most of these proteins were distributed among three categories related to Transcription (category K: 307), Amino acid transport and metabolism (category E: 222) and Carbohydrate transport and metabolism (category G: 268) (Fig. 3).

The genome was screened using various antimicrobial gene databases, including NCBI Bacterial Antimicrobial Resistance Finder Plus, ResFinder 4.5, and Comprehensive Antibiotic Resistance Database [4]. No antimicrobial resistance genes or virulent genes were identified. Further investigation of the PlasmidFinder server revealed the presence of plasmid pR18 (assessment number JN601038) in the genome (Supplementary Fig. 1). Studies have shown that this plasmid has lincomycin resistance, can be transferred from one bacterium to another, and is one of the most common plasmids in the genera *Bacillus* and *Lactobacillus* [5]. The draft genome sequence dataset was deposited at the NCBI Genebank with accession number JARUWW000000000.

Based on the data presented above, the strain 33C was unequivocally identified as *L. plantarum* 33C. In addition, the safety-related data described, confirm that the strain *L. plantarum* 33C is safe and does not possess antimicrobial resistance genes or virulent genes.

4. Experimental Design, Materials and Methods

4.1. DNA extraction and whole genome sequencing

L. plantarum 33C was grown overnight in MRS broth (Oxoid, Wesel, Germany) at 35 °C. The QIAGEN DNeasy PowerSoil Pro Kit was used to isolate DNA from the bacteria following the
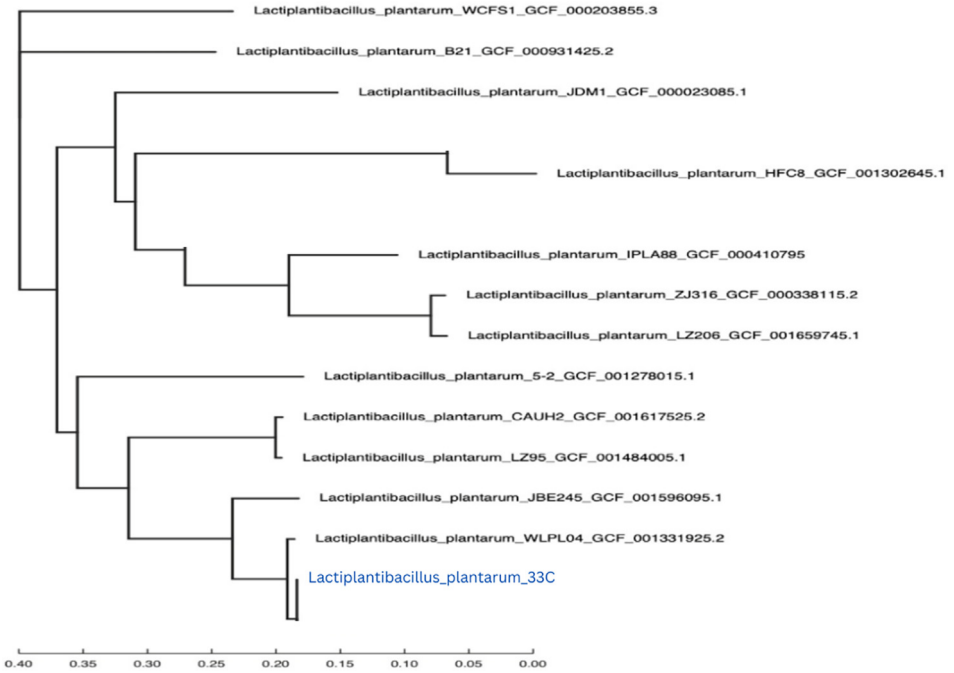
Fig. 2. Phylogenetic tree of *L. plantarum* 33C. FastTree2 was utilized for reconstructing the phylogenetic relationship using the core-genome SNPs. The GenBank accession numbers are listed in the phylogenetic tree.
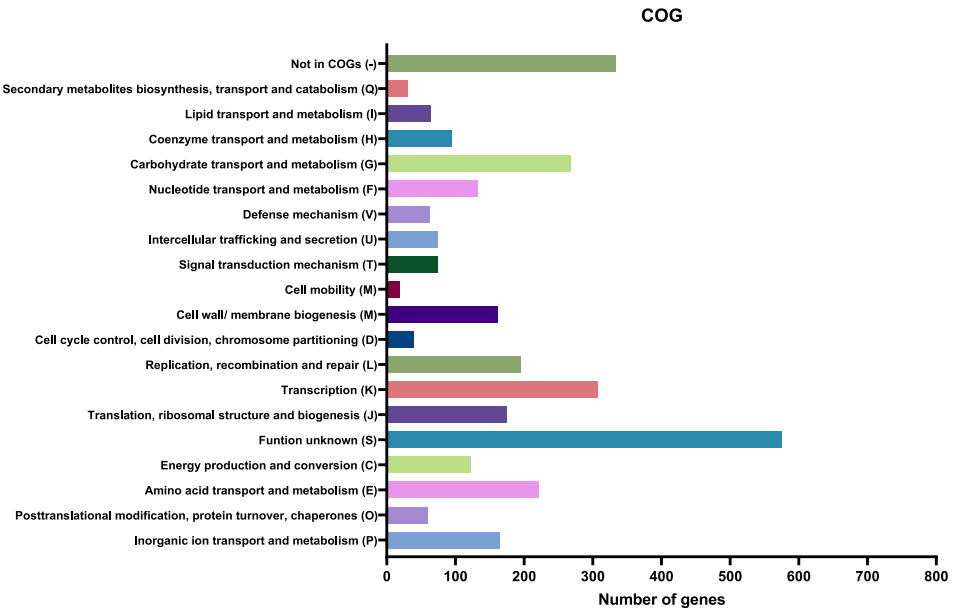

Fig. 3. Distribution of *L. plantarum* 33C proteins in COG Categories.

manufacturer's protocol. Briefly, the quantity of DNA in the samples was measured using the QuantiFluor® dsDNA System with the GloMax Plate Reader System (Promega Italia, Promega France). To prepare the DNA libraries, Nextera XT DNA Library Preparation Kit (Illumina, Warszawa, Poland) was used coupled with IDT Unique Dual Indexes. An amount of 1 ng DNA was used as input. The genomic DNA was fragmented using Illumina Nextera XT fragmentation enzyme and a unique dual index was added to the sample. The library was then constructed using 12 cycles of PCR. The DNA library was purified using AMpure magnetic beads and eluted in QIAGEN EB buffer. The quantity of DNA in the library was measured with a Qubit 4 fluorometer and a Qubit dsDNA HS Assay Kit. The library was then sequenced with the Illumina NovaSeq platform with 2 × 150 bp reads. The CG view server constructed a circular genomic map from the resultant genome (https://cgview.ca/) [6].

4.2. Taxonomic identification of the strain

The raw paired-end reads underwent trimming and processing using BBDuk with a read quality trimming parameter of 22. SPAdes were then utilized to assemble the fastqs with the "–careful" parameter. CheckM's lineage_wf function was used to evaluate the completeness of the assembly. The CosmosID core genome SNP typing pipeline was employed to analyze the assembled contigs and make epidemiological conclusions based on phylogenetic placement and SNP variations. Parsnp served as the core genome aligner for mapping the core genome of multiple microbial genomes. The final core-genome SNPs were used to reconstruct the phylogenomic relationship among the genomes using FastTree2, as depicted in Fig. 2.

4.3. Functional annotation and gene prediction

Genome annotation was carried out using the Prokaryotic Genome Annotation System (Prokka) and the Rapid Annotations using Subsystems Technology (RAST) webserver (https://rast.nmpdr.org/) [7]. The Cluster of Orthologous groups (COG) annotations were generated using eggNOG-mapper, based on the eggNOG v.4.5 orthology database [8].

4.4. Search for antimicrobial resistance genes and virulence factors

For antimicrobial resistance genes (AMR) and virulence factor (VF) detection, the assembled genome of PN39MY was compared to the Resfinder AMR and VFDB VF database [9,10] using ABRicate version 1.0.1. AMR and VF genes [11] were considered present only if the sequences matched the assembled genome at a nucleotide identity >90 % and the alignment coverage of the gene's sequence length was >60 %. The average nucleotide identity (ANIm) between isolates was calculated using MUMmer. Prokka Annotation Pipeline was then used to annotate the genome.

4.5. Screening for plasmids

PlasmidFinder database [12] (software version 2.0.1), threshold identity set to 95 %, minimum coverage 60 % and Blast searches were used to search for plasmid-related contigs in the sequenced genome, the circular contigs presence was examined in the assembly files.

Limitations

None.

Ethics Statement

This study did not involve any human subjects or animal experiments. No ethical approval was required.

CRediT Author Statement

Ashwinipriyadarshini Megur: Data curation, Software, Writing – original draft, Methodology, Formal analysis, Writing – review & editing; **Eglė Lastauskienė**: Supervision, Project administration, Manuscript review & editing; **Aurelijus Burokas**: Supervision, Project administration, Manuscript review & editing.

Data Availability

In vitro screening and characterization of lactic acid bacteria from Lithuanian fermented food with potential probiotic properties (Reference data) (National Center for Biotechnology Information (NCBI)).

Acknowledgments

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.

Supplementary Materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.dib.2024.110750.

References

[1] A.G. Tóth, M.F. Judge, S.Á. Nagy, M. Papp, N. Solymosi, A survey on antimicrobial resistance genes of frequently used probiotic bacteria, 1901 to 2022, Eurosurveillance 28 (2023) 1901–2022, doi:10.2807/1560-7917.ES.2023.28.14.2200272.

[2] L.G. Pell, R.G. Horne, S. Huntley, H. Rahman, S. Kar, M.S. Islam, K.C. Evans, S.K. Saha, A. Campigotto, S.K. Morris, D.E. Roth, P.M. Sherman, Antimicrobial susceptibilities and comparative whole genome analysis of two isolates of the probiotic bacterium Lactiplantibacillus plantarum, strain ATCC 202195, Sci. Rep. 11 (2021) 15893, doi:10.1038/s41598-021-94997-6.

[3] A. Megur, E.B.M. Daliri, T. Balnionytė, J. Stankevičiūtė, E. Lastauskienė, A. Burokas, In vitro screening and characterization of lactic acid bacteria from Lithuanian fermented food with potential probiotic properties, Front. Microbiol. 14 (2023) 1–18, doi:10.3389/fmicb.2023.1213370.

[4] B.P. Alcock, W. Huynh, R. Chalil, K.W. Smith, A.R. Raphenya, M.A. Wlodarski, A. Edalatmand, A. Petkau, S.A. Syed, K.K. Tsang, S.J.C. Baker, M. Dave, M.C. Mccarthy, K.M. Mukiri, J.A. Nasir, B. Golbon, H. Imtiaz, X. Jiang, K. Kaur, M. Kwong, Z.C. Liang, K.C. Niu, P. Shan, J.Y.J. Yang, K.L. Gray, G.R. Hoad, B. Jia, T. Bhando, L.A. Carfrae, M.A. Farha, S. French, R. Gordzevich, K. Rachwalski, M.M. Tu, E. Bordeleau, D. Dooley, E. Griffiths, H.L. Zubyk, E.D. Brown, F. Maguire, R.G. Beiko, W.W.L. Hsiao, F.S.L. Brinkman, G. Van Domselaar, A.G. Mcarthur, CARD 2023: expanded curation, support for machine learning, and resistome prediction at the comprehensive antibiotic resistance database, Nucleic Acids Res. 51 (2023) D690–D699, doi:10.1093/nar/gkac920.

[5] T. Jalilsood, A. Baradaran, F.H. Ling, S. Mustafa, K. Yusof, R.A. Rahim, Characterization of pR18, a novel rolling-circle replication plasmid from Lactobacillus plantarum, Plasmid 73 (2014) 1–9, doi:10.1016/j.plasmid.2014.04.004.

[6] J.R. Grant, P. Stothard, The CGView Server: a comparative genomics tool for circular genomes, Nucleic Acids Res. 36 (2008), doi:10.1093/nar/gkn179.

[7] R.K. Aziz, D. Bartels, A. Best, M. DeJongh, T. Disz, R.A. Edwards, K. Formsma, S. Gerdes, E.M. Glass, M. Kubal, F. Meyer, G.J. Olsen, R. Olson, A.L. Osterman, R.A. Overbeek, L.K. McNeil, D. Paarmann, T. Paczian, B. Parrello, G.D. Pusch, C. Reich, R. Stevens, O. Vassieva, V. Vonstein, A. Wilke, O. Zagnitko, The RAST Server, Rapid annotations using subsystems technology, BMC Genomics 9 (2008), doi:10.1186/1471-2164-9-75.

[8] J. Huerta-Cepas, K. Forslund, L.P. Coelho, D. Szklarczyk, L.J. Jensen, C. Von Mering, P. Bork, Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper, Mol. Biol. Evol. 34 (2017) 2115–2122, doi:10.1093/molbev/msx148.

[9] L. Chen, J. Yang, J. Yu, Z. Yao, L. Sun, Y. Shen, Q. Jin, VFDB: a reference database for bacterial virulence factors, Nucleic Acids Res. 33 (2005), doi:10.1093/nar/gki008.

[10] M. Feldgarden, V. Brover, D.H. Haft, A.B. Prasad, D.J. Slotta, I. Tolstoy, G.H. Tyson, S. Zhao, C.H. Hsu, P.F. McDermott, D.A. Tadesse, C. Morales, M. Simmons, G. Tillman, J. Wasilenko, J.P. Folster, W. Klimke, Validating the AMRFINder tool and resistance gene database by using antimicrobial resistance genotype-phenotype correlations in a collection of isolates, Antimicrob. Agents Chemother. 63 (2019), doi:10.1128/AAC.00483-19.

[11] L. Chen, D. Zheng, B. Liu, J. Yang, Q. Jin, VFDB 2016: hierarchical and refined dataset for big data analysis - 10 years on, Nucleic Acids Res. 44 (2016) D694–D697, doi:10.1093/nar/gkv1239.

[12] A. Carattoli, E. Zankari, A. Garciá-Fernández, M.V. Larsen, O. Lund, L. Villa, F.M. Aarestrup, H. Hasman, In silico detection and typing of plasmids using plasmidfinder and plasmid multilocus sequence typing, Antimicrob. Agents Chemother. 58 (2014) 3895–3903, doi:10.1128/AAC.02412-14.