

JAST (Journal of Animal Science and Technology) TITLE PAGE

Upload this completed form to website with submission

ARTICLE INFORMATION	Fill in information in each box below
Article Type	Research Article
Article Title (within 20 words without abbreviations)	Complete genome sequence of <i>Lactococcus taiwanensis</i> strain K_LL004, encoding hydrolytic enzymes of plant polysaccharides isolated from grasshopper (<i>Oxya chinensis sinuosa</i>)
Running Title (within 10 words)	Complete genome sequence of <i>Lactococcus taiwanensis</i> strain K_LL004
Author	Hyunok Doo ^{1#} , Hyeri Kim ^{1#} , Jin Ho Cho ^{2#} , Minho Song ^{3#} , Eun Sol Kim ¹ , Jae Hyoung Cho ¹ , Sheena Kim ¹ , Gi Beom Keum ¹ , Jinok Kwak ¹ , Srinivas Pandey ¹ , Hyeun Bum Kim ^{1*} , and Ju-Hoon Lee ^{4*}
Affiliation	1 Department of Animal Resources Science, Dankook University, Cheonan, South Korea 2 Division of Food and Animal Science, Chungbuk National University, Cheongju, South Korea 3 Division of Animal and Dairy Science, Chungnam National University, Daejeon, South Korea 4 Department of Food Animal Biotechnology, Department of Agricultural Biotechnology, Center for Food and Bioconvergence, Seoul National University, Seoul, South Korea
ORCID (for more information, please visit https://orcid.org)	Hyunok Doo: https://orcid.org/0000-0003-4329-4128 Hyeri Kim: https://orcid.org/0000-0002-6560-2390 Jin Ho Cho: https://orcid.org/0000-0001-7151-0778 Minho Song: https://orcid.org/0000-0002-4515-5212 Eun Sol Kim: https://orcid.org/0000-0001-8801-421X Jae Hyoung Cho: https://orcid.org/0000-0002-1128-3451 Sheena Kim: https://orcid.org/0000-0002-5410-1347 Gi Beom Keum: https://orcid.org/0000-0001-6006-9577 Jinok Kwak: https://orcid.org/0000-0003-1217-3569 Srinivas Pandey: https://orcid.org/0000-0002-6947-3469 Hyeun Bum Kim: https://orcid.org/0000-0003-1366-6090 Ju-Hoon Lee: https://orcid.org/0000-0003-0405-7621
Competing interests	No potential conflict of interest relevant to this article was reported.
Funding sources State funding sources (grants, funding sources, equipment, and supplies). Include name and number of grant if available.	This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education (2022M3A9I5082342 & 2021R1I1A3059910).
Acknowledgements	Not applicable.
Availability of data and material	The complete genome sequences of <i>Lactococcus taiwanensis</i> K_LL004 were deposited in GeneBank under the accession numbers CP070872.1. The BioSample accession number is SAMN17981207, and BioProject accession number is PRJNA224116
Authors' contributions Please specify the authors' role using this form.	Conceptualization: Doo H, Kim HB, Lee JH. Formal analysis: Kim ES, Cho JH, Kim S, Keum GB. Methodology: Cho JH, Song M. Validation: Kim S, Kwak J, Pandey S. Writing - original draft: Doo H, Kim H, Cho JH, Song M. Writing - review & editing: Pandey S, Kim H, Kim HB, Lee JH.
Ethics approval and consent to participate	This article does not require IRB/IACUC approval because there are no human and animal participants.

6 **CORRESPONDING AUTHOR CONTACT INFORMATION**

For the corresponding author (responsible for correspondence, proofreading, and reprints)	Fill in information in each box below
First name, middle initial, last name	Hyeun Bum Kim
Email address – this is where your proofs will be sent	hbkim@dankook.ac.kr
Secondary Email address	
Address	Department of Animal Resources Science, Dankook University, Cheonan, South Korea
Cell phone number	+82-10-3724-3416
Office phone number	+82-41-550-3653
Fax number	+82-41-565-2940

7
8 **CORRESPONDING AUTHOR CONTACT INFORMATION**

For the corresponding author (responsible for correspondence, proofreading, and reprints)	Fill in information in each box below
First name, middle initial, last name	Ju-Hoon Lee
Email address – this is where your proofs will be sent	juhlee@snu.ac.kr
Secondary Email address	
Address	Department of Food Animal Biotechnology, Department of Agricultural Biotechnology, Center for Food and Bioconvergence, Seoul National University, Seoul, South Korea
Cell phone number	+82-10- 9678-5529
Office phone number	+82-2-880-4854
Fax number	+82-2-875-5095

9
10
11

Complete genome sequence of *Lactococcus taiwanensis* strain K_LL004, encoding hydrolytic enzymes of plant polysaccharides isolated from grasshopper (*Oxya chinensis sinuosa*)

Hyunok Doo^{1#}, Hyeri Kim^{1#}, Jin Ho Cho^{2#}, Minho Song^{3#}, Eun Sol Kim¹, Jae Hyoung Cho¹, Sheena Kim¹, Gi Beom Keum¹, Jinok Kwak¹, Srinivas Pandey¹, Hyeun Bum Kim^{1*} and Ju-Hoon Lee^{4*}

¹ Department of Animal Resources Science, Dankook University, Cheonan, South Korea

² Division of Food and Animal Science, Chungbuk National University, Cheongju, South Korea

³ Division of Animal and Dairy Science, Chungnam National University, Daejeon, South Korea

⁴ Department of Food Animal Biotechnology, Department of Agricultural Biotechnology, Center for Food and Bioconvergence, Seoul National University, Seoul, South Korea

Equal contributors

* Corresponding authors

Hyeun Bum Kim

Department of Animal Resources Science, Dankook University, Cheonan, South Korea

Tel: +82-41-550-3653

Email: hbkim@dankook.ac.kr

Ju-Hoon Lee

Department of agricultural biotechnology. Seoul National University, Seoul, South Korea

Tel: +82-2-880-4854

Email: juhlee@snu.ac.kr

Abstract

The *Lactococcus taiwanensis* strain K_LL004 was isolated from the gut of a grasshopper (*Oxya chinensis sinuosa*) collected from local farm in Korea. *L. taiwanensis* strain K_LL004 is the functional probiotic candidate with an ability to hydrolyse plant polysaccharides. The complete genome of the *L. taiwanensis* strain K_LL004 contains one circular chromosome (1,995,099 bp) with a guanine + cytosine (GC) content of 38.8%. Moreover, 1,929 Protein-coding sequence, 19 rRNA genes, and 62 tRNA genes were identified based on results of annotation. *L. taiwanensis* strain K_LL004 has a gene, which encodes hydrolytic enzymes such as beta-glucosidase and beta-xylosidase, that hydrolyzes plant polysaccharides.

Keywords (3 to 6):

Lactococcus taiwanensis, grasshopper, beta-glucosidase, beta-xylosidase, Whole genome sequencing

Lactococcus is a genus of lactic acid bacteria (LAB) that are present on grass and other plant material and in the gastrointestinal tracts[1]. Twenty-two species of the genus *Lactococcus* are established till date. In particular, *Lactococcus lactis* is the most common strain which is used as a starter in food fermentation [2]. *Lactococcus taiwanensis*, a type of Lactic acid Bacteria, has not been studied in detail, and therefore the genomic information of *Lactococcus taiwanensis* is limited.

In the present study, the *L. taiwanensis* strain K_LL004 was isolated from the gut of a grasshopper (*Oxya chinensis sinuosa*), an insect preferring to feed on plants, collected from local farm in Yangyang, Gangwon-do, Korea. The *L. taiwanensis* strain K_LL004 was grown in de Man-Rogosa-Sharpe broth at 37°C for 24 h. Genomic DNA was extracted using the MagAttract HMW DNA Kit (QIAGEN, Hilden, Germany), according to the manufacturer's instructions. The complete genome of the *L. taiwanensis* strain K_LL004 was sequenced using the PacBio RS II (Pacific Biosciences, Menlo Park, CA, USA) platform at Insilicogen (Yongin, Korea). Library preparation was performed using SMRTbell™ Template Prep Kit 1.0 following the manufacturer's instructions (Pacific Biosciences). PacBio sequencing produced 161,058 of long reads and 1,143,521,995 base pairs after subreads filtering. *De novo* assembly was performed using the hierarchical genome assembly process (HGAP v2.3.0) workflow and polished using Quiver. The quality of genome assembly was assessed by using Quality Assessment Tool for Genome Assemblies (QUAST) v5.0.2 [3]. The quantitative assessment of the genome completeness was conducted by using Benchmarking Universal Single-Copy Orthologs (BUSCO) v3.0.2 [4]. Protein coding genes, rRNA and tRNA genes of *L. taiwanensis* strain K_LL004 were functionally annotated and predicted through Rapid Annotation using Subsystem Technology (RAST) v2.0 [5]. The functional categorization of all predicted Protein coding genes was performed using Clusters of Orthologous Groups (COG)-based EggNOG-mapper v2 [6]. Potential virulence factors and antibiotic resistance in *L. taiwanensis* strain K_LL004 were predicted using the BLASTn method according to the Virulence Factor Database (VFDB) and the Comprehensive Antibiotic Resistance Database (CARD) [7, 8].

The complete genome of the *L. taiwanensis* strain K_LL004 contains one circular chromosome (1,995,099 bp) with a guanine + cytosine (GC) content of 38.8%, 1,929 predicted protein-coding sequence, 19 rRNA genes, and 62 tRNA genes. The genome feature and map of *L. taiwanensis* strain K_LL004 are illustrated in Table 1, Fig 1A, and Fig 1B.

84 It was confirmed that the *L. taiwanensis* strain K_LL004 has genes which encodes
85 enzymes like beta-glucosidase (EC 3.2.1.21 BG) and beta-xylosidase (EC 3.2.1.37 xyl3), which
86 plays an important role in beta-glycoside metabolism and xylose utilization, respectively. These
87 enzymes are known to hydrolyze the plant cell wall polysaccharides [9]. In addition, the genome
88 of *L. taiwanensis* strain K_LL004 didn't show presence of any virulence factors and antibiotic
89 resistant genes, indicating that *L. taiwanensis* strain K_LL004 can be speculated as a potential
90 probiotic candidate with an ability to hydrolyse plant polysaccharides.

91 Acknowledgments

92 Not applicable.

93

ACCEPTED

References (Vancouver or NLM style)

1. McAuliffe O. Symposium review: *Lactococcus lactis* from nondairy sources: Their genetic and metabolic diversity and potential applications in cheese. *J Dairy Sci.* 2018;101(4):3597-610;https://doi.org/10.3168/jds.2017-13331
2. Lopez-Gonzalez MJ, Escobedo S, Rodriguez A, Neves AR, Janzen T, Martinez B. Adaptive Evolution of Industrial *Lactococcus lactis* Under Cell Envelope Stress Provides Phenotypic Diversity. *Front Microbiol.* 2018;9:2654;https://doi.org/10.3389/fmicb.2018.02654
3. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics.* 2013;29(8):1072-5;https://doi.org/10.1093/bioinformatics/btt086
4. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics.* 2015;31(19):3210-2;https://doi.org/10.1093/bioinformatics/btv351
5. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics.* 2008;9:75;https://doi.org/10.1186/1471-2164-9-75
6. Cantalapiedra CP, Hernandez-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Mol Biol Evol.* 2021;38(12):5825-9;https://doi.org/10.1093/molbev/msab293
7. Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, et al. VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Res.* 2005;33(Database issue):D325-8;https://doi.org/10.1093/nar/gki008
8. McArthur AG, Wagglechner N, Nizam F, Yan A, Azad MA, Baylay AJ, et al. The comprehensive antibiotic resistance database. *Antimicrob Agents Chemother.* 2013;57(7):3348-57;https://doi.org/10.1128/AAC.00419-13
9. Kaupert Neto AA, Borin GP, Goldman GH, Damasio AR, Oliveira JV. Insights into the plant polysaccharide degradation potential of the xylanolytic yeast *Pseudozyma brasiliensis*. *FEMS Yeast Res.* 2016;16(2):fov117;https://doi.org/10.1093/femsyr/fov117

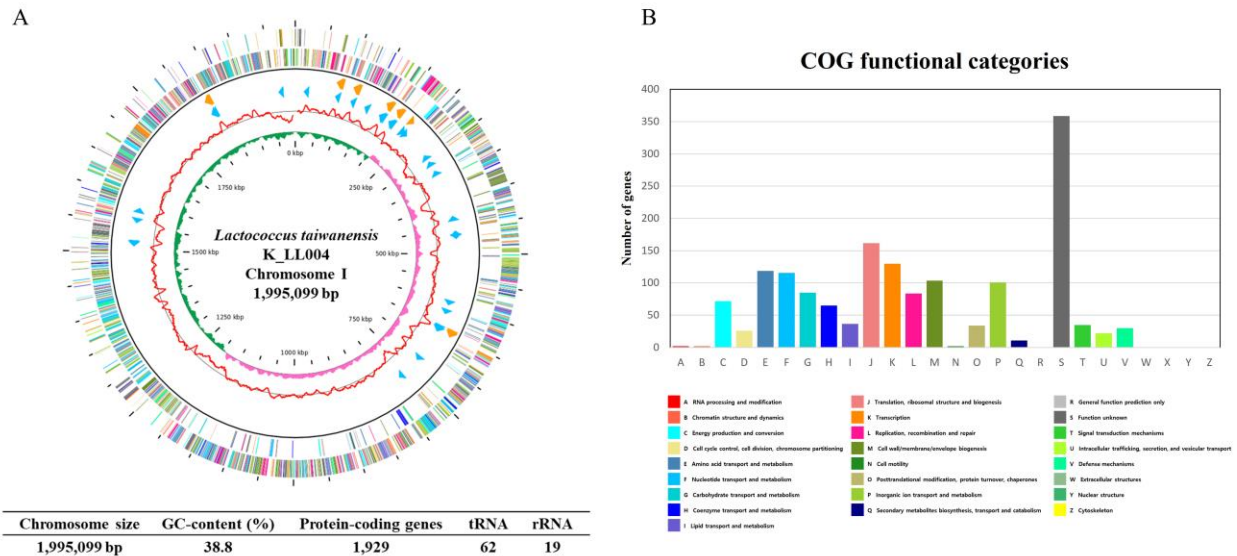


Fig 1. Genome map of *Lactococcus taiwanensis* K_LL004 and the functional categorization of predicted protein coding genes.

The outer circle denotes the locations of all annotated open reading frames (ORFs), and the inner circle with the red denotes guanine + cytosine (GC) content. Pink, and green peaks denote GC skew. The orange, and sky-blue arrows denote the rRNAs, and tRNA operons, respectively. The annotated ORFs are colored differently based on the Clusters of Orthologous Groups (COG) assignments (Fig 1A). COG functional categories of predicted protein coding genes (Fig 1B).

Table 1. Genome features of *Lactococcus taiwanensis* strain K_LL004

Property	Term
Average genome coverage	449x
Chromosome length (bp)	1,995,099 bp
No. of contig	1 (chromosome)
Guanine + cytosine (G + C) content (%)	38.8
Protein-coding genes	1,929
rRNA genes	19
tRNA genes	62
Genbank Accession No.	CP070872