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)	Genome analysis of *Lactococcus taiwanensis* strain K_LL001 with potential cellulose degrading functions
Running Title (within 10 words)	Genome analysis of *Lactococcus taiwanensis* strain K_LL001
Author	Eun Sol Kim[1#], Jin Ho Cho[2#], Minho Song[3#], Sheena Kim[1], Gi Beom Keum[1], Hyunok Doo[1], Jinok Kwak[1], Sriniwas Pandey[1], Sumin Ryu[1], Yejin Choi[1], Juyoun Kang[1], Hyeun Bum Kim[1*] and Ju-Hoon Lee[4*]
Affiliation	[1] Department of Animal Resources Science, Dankook University, Cheonan 31116, Korea [2] Division of Food and Animal Science, Chungbuk National University, Cheongju 28644, Korea [3] Division of Animal and Dairy Science, Chungnam National University, Daejeon 34134, Korea [4] Department of Food Animal Biotechnology, Department of Agricultural Biotechnology, Center for Food and Bioconvergence, Seoul National University, Seoul 08826, Korea
ORCID (for more information, please visit https://orcid.org)	Eun Sol Kim: https://orcid.org/0000-0001-8801-421X Minho Song: https://orcid.org/0000-0002-4515-5212 Jin Ho Cho: https://orcid.org/0000-0001-7151-0778 Sheena Kim: https://orcid.org/0000-0002-5410-1347 Gi Beom Keum: https://orcid.org/0000-0001-6006-9577 Hyunok Doo: https://orcid.org/0000-0003-4329-4128 Jinok Kwak: https://orcid.org/0000-0003-1217-3569 Sriniwas Pandey: https://orcid.org/0000-0002-6947-3469 Sumin Ryu: https://orcid.org/0000-0002-1569-3394 Yejin Choi: https://orcid.org/0000-0002-7434-299X Juyoun Kang: https://orcid.org/0000-0002-3974-2832 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(NRF-2021R1I1A3059910 and NRF-2019M3A9F3065227).
Acknowledgements	Not applicable.
Availability of data and material	The complete genome sequences of of *Lactococcus taiwanensis* strain K_LL001 were deposited in GenBank under the accession numbers NZ_CP070381. The BioSample accession number is SAMN17922033, and BioProject accession number is PRJNA702013
Authors' contributions Please specify the authors' role using this form.	Conceptualization: Kim HB, Lee JH Data curation: Kim ES, Kim HB Formal analysis: Kim ES Methodology: Keum GB, Kwak J, Pandey S, Ryu SM Software: Doo H, Choi Y, Kang J Validation: Cho JH, Song M, Kim S, Kim HB Writing - original draft: Kim ES Writing - review & editing: Kim ES, Cho JH, Song M, Kim S, Keum GB, Doo H, Kwak J, Pandey S, Ryu SM, Choi Y, Kang J, 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.

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
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 08826, Korea
Cell phone number	+82-10-9678-5529
Office phone number	+82-2-880-4854
Fax number	+82-2-875-5095

Genome analysis of *Lactococcus taiwanensis* strain K_LL001 with potential cellulose degrading functions.

Eun Sol Kim[1#], Jin Ho Cho[2#], Minho Song[3#], Sheena Kim[1], Gi Beom Keum[1], Hyunok Doo[1], Jinok Kwak[1], Sriniwas Pandey[1], Sumin Ryu[1], Yejin Choi[1], Juyoun Kang[1], Hyeun Bum Kim[1*] and Ju-Hoon Lee[4*]

[1] Department of Animal Resources Science, Dankook University, Cheonan 31116, Korea

[2] Division of Food and Animal Science, Chungbuk National University, Cheongju 28644, Korea

[3] Division of Animal and Dairy Science, Chungnam National University, Daejeon 34134, Korea

[4] Department of Food Animal Biotechnology, Department of Agricultural Biotechnology, Center for Food and Bioconvergence, Seoul National University, Seoul 08826, 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 08826, Korea

Tel: +82-2-880-4854

Email: juhlee@snu.ac.kr

Abstract

Lactococcus taiwanensis starin K_LL001 was isolated from the gut of grasshopper (*Oxya chinensis sinuosa*). In this study, we presented the complete genome sequence of *L. taiwanensis* stain K_LL001. The genome of K_LL001 genome was composed of 1 circular chromosome without plasmids. The length of the whole genome was 2,018,259bp, guanin + cytosine (G±C) content (%) was 38.75%, with 2,021 predicted protein-coding sequences (CDS). The most abundant CAZyme class in *L. taiwanensis* strain K_LL001 was glycoside hydrolases (GH) class. GHs is the key enzymes involved in carbohydrate metabolism, and they catalyze the hydrolysis of glycosidic bonds in complex carbohydrates such as cellulose, hemicellulose, and starch. Moreover, *L. taiwanensis* strain K_LL001 has genes encoding enzymes which can catalyze the transformation of one glycoside to another. Overall, this study will contribute to a further understanding of *L. taiwanensis* strain K_LL001 at the genomic level and provide a theoretical basis for its future application in swine industry.

Keywords (3 to 6):

Lactococcus taiwanensis, pig, grasshopper, glycoside hydrolases, carbohydrates

Main text

Beneficial microorganisms can colonize the host's intestinal tract and offer benefits to the host due to their unique abilities such as the production of digestive enzymes that enhance feed digestion and absorption [1]. Complex carbohydrate, such as cellulose, cannot be digested by pigs but can only be metabolized by the swine gut microbiota, serving as an important energy source for pigs. Since a significant quantity of cellulose is present in nursery and finisher pig feed, it is desirable to enhance its utilization for improved energy efficiency. As a result, there is a growing interest in researching bacteria associated with cellulose utilization in the swine industry [2], given the substantial cellulose content in swine feed. Insects, such as termites (*Isoptera*), bookworms (*Lepidoptera*), and others have been found to harbor symbiotic microflora in their guts, which is responsible for the digestion of cellulosic feed [3]. In this study, we isolated *Lactococcus taiwanensis* from the gut of a grasshopper (*Oxya chinensis sinuosa*). It exhibited a low DNA sequence similarity with *Lactococcus lactis* spp.[4]. Due to its recent discovery, there is relatively limited genomic information available for *L. taiwanensis.* The aim of this study was to contribute to a more comprehensive understanding of *L. taiwanensis* at the genomic level.

L. taiwanensis strain K_LL001 was isolated from the gut of a grasshopper (*Oxya chinensis sinuosa*), collected from the local grasshopper farm in Yangyang, Gangwon-do, Korea. The K_LL001 was grown in MRS broth (BD Difco™, New Jersey, USA) at 37°C for 24hours. 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_LL001 was sequenced using the PacBio RS II (Pacific Biosciences, Menlo Park, USA) platform at Insilicogen (Yongin, Korea). Library preparation was performed using SMRTbell™ Template Prep Kit 1.0 following the manufacturer's instructions (Pacific Biosciences, Menlo Park, USA). A total number of 139,220 long read sequences (854,450,914 base pairs) were produced after subreads filtering sequences. De novo assembly of the gene sequences were performed using the hierarchical genome assembly process (HGAP v2.3.0) workflow, and further polished with Quiver. Quality Assessment Tool for Genome Assemblies (QUAST) v5.2.0 was used for evaluating genome assembly [5]. Benchmarking Universal Single-Copy Orthologs (BUSCO) v5.4.7 was used for assessing genome completeness and contamination [6]. Prediction of protein coding genes, rRNA and tRNA genes were identified through Rapid Annotation using Subsystem Technology (RAST) server v2.0. Clusters of Orthologous Groups (COGs)-based EggNOG-mapper v2.0 was used to predict functional categorization of protein coding genes [7]. Virulence Factor Database (VFDB) and the Comprehensive Antibiotic Resistance Database (CARD) was used to predict potential

virulence factors and antibiotic resistances [8, 9]. For the identification of Carbohydrate-Active enzyme (CAZyme), data were submitted to automated carbohydrate-active enzyme and substrate annotation (dbCAN3) website [10].

The complete genome of the *L. taiwanensis* strain K_LL001 contained one circular chromosome (2,018,259 bp) with a guanine + cytosine (GC) content of 38.75% and no plasmid was found. A total of 2,021 predicted protein-coding sequences, 19 rRNA genes, and 60 tRNA genes were identified in *L. taiwanensis* strain K_LL001. *L. taiwanensis* strain K_LL001 had no virulence factor and antibiotic resistance gene. The genome features and genome map of *L. taiwanensis* strain K_LL001 were illustrated in **Figure 1.**

All predicted proteins were subjected to the COG database for functional classification assignment. Top four classifications except S (uncharacterized genes) were J, K, E, and F. The genes categorized under Category J and Category K are associated with translation and transcription of genes in the bacterium. Category E (Amino acid transport and metabolism) and Category F (Nucleotide transport and metabolism) are involved in transport and metabolize amino acids and nucleotides, which are key for bacterial growth and survival.

The summarized distribution of predicted CAZyme was as follows: auxiliary activity (AA) 39; carbohydrate esterases (CE) 34, glycoside hydrolases (GH) 587, glycosyl transferases (GT) 189, carbohydrate-binding module (CBM) 22, and polysaccharide lyases (PL) 2. The CAZyme class mostly possessed in *L. taiwanensis* strain K_LL001 was GH class. According to Architecture et Fonction des Macromolécules Biologiques (AFMB) laboratory, GHs are the key enzymes involved in carbohydrate metabolism, and they catalyze the hydrolysis of glycosidic bonds in complex carbohydrates such as cellulose, hemicellulose, and starch. In addition, it was verified that *L. taiwanensis* strain K_LL001 had enzymes which can catalyze the transformation of one glycoside to another. For instance, glycosyltransferase(EC 2.4) is enzyme that catalyze the formation of the glycosidic linkage to form a glycoside, and GH92 family glycosyl hydrolase (EC 3.2.1) genes are a group of glycosyl hydrolases that catalyze the hydrolysis of specific glycosidic bonds in carbohydrates. Moreover, *L. taiwanensis* strain K_LL001 had no virulence factors or antibiotic-resistant genes. Overall, the genomic characteristics of *L. taiwanensis* strain K_LL001 suggest that it could be used as probiotics to increase swine performance through enhanced carbohydrate utilization in feed.

Acknowledgments

Not applicable.

Nucleotide sequence accession number

The complete genome sequences of of *Lactococcus taiwanensis* strain K_LL001 were deposited in GenBank under the accession numbers NZ_CP070381. The BioSample accession number is SAMN17922033, and BioProject accession number is PRJNA702013.

References

1. Li F, Xie Y, Gao X, Shan M, Sun C, Niu YD, et al. Screening of cellulose degradation bacteria from Min pigs and optimization of its cellulase production. Electronic Journal of Biotechnology. 2020;48:29-35;https://doi.org/10.1016/j.ejbt.2020.09.001.

2. Varel V. Activity of fiber-degrading microorganisms in the pig large intestine. Journal of animal science. 1987;65(2):488-96;https://doi.org/10.2527/jas1987.652488x.

3. Gupta P, Samant K, Sahu A. Isolation of cellulose-degrading bacteria and determination of their cellulolytic potential. International journal of microbiology. 2012;2012; https://doi.org/10.1155/2012/578925.

4. Chen Y-s, Chang C-h, Pan S-f, Wang L-t, Chang Y-c, Wu H-c, et al. Lactococcus taiwanensis sp. nov., a lactic acid bacterium isolated from fresh cummingcordia. International journal of systematic and evolutionary microbiology. 2013;63(Pt_7):2405-9;https://doi.org/10.1099/ijs.0.045757-0.

5. 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.

6. Seppey M, Manni M, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness. Gene prediction: methods and protocols. 2019:227-45;https://doi.org/10.1093/bioinformatics/btv351.

7. 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(1):1-15;https://doi.org/10.1186/1471-2164-9-75.

8. Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, et al. VFDB: a reference database for bacterial virulence factors. Nucleic acids research. 2005;33(suppl_1):D325-D8;https://doi.org/10.1093/nar/gki008.

9. Alcock BP, Raphenya AR, Lau TT, Tsang KK, Bouchard M, Edalatmand A, et al. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic acids research. 2020;48(D1):D517-D25;https://doi.org/10.1093/nar/gkz935.

10. Ausland C, Zheng J, Yi H, Yang B, Li T, Feng X, et al. dbCAN-PUL: a database of experimentally characterized CAZyme gene clusters and their substrates. Nucleic Acids Research. 2021;49(D1):D523-D8;https://doi.org/10.1093/nar/gkaa742.

Table

Table 1. Genome features of *Lactococcus taiwanensis* strain K_LL001

Property	Term
Library used	PacBio SMRTbell™ library
Sequencing platforms	PacBio RS II sequencer
Average genome coverage	327x
Chromosome length (bp)	2,018,259 bp
No. of contigs	1
G + C content (%)	38.75%
Protein-coding genes	2,021
rRNA genes	19
tRNA genes	60
GenBank Accession No.	SAMN17922033

Figure

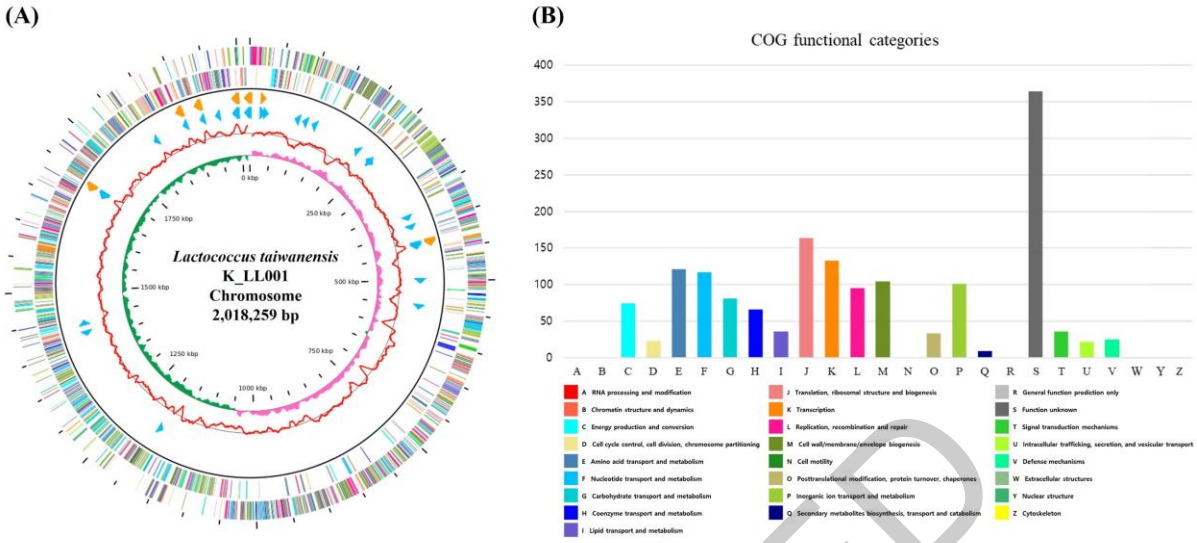

Figure 1. Genome map of *Lactococcus taiwanensis* K_LL001. (A) From the outside to the inside: the outer circle indicates all annotated open reading frames (ORFs), while the inner circle in red represents guanine + cytosine (GC) content. The annotated ORFs are color-coded based on Clusters of Orthologous Groups (COG) assignments, and rRNAs and tRNA operons are denoted by orange and sky-blue arrows. The inner circle indicates GC skew by pink and green peaks, respectively. (B) Functional categories of predicted protein-coding genes according to COG.