

Genome sequence analysis of potential probiotic strain *Leuconostoc lactis* EFEL005 isolated from kimchi

Jin Seok Moon¹, Hye Sun Choi², So Yeon Shin¹,
Sol Ji Noh¹, Che Ok Jeon³, and Nam Soo Han^{1*}

¹Brain Korea 21 Center for Bio-Resource Development, Division of Animal, Horticultural, and Food Sciences, Chungbuk National University, Cheongju 362-763, Republic of Korea

²Department of Agro-food Resource, National Academy of Agricultural Science, RDA, Jeonju 560-500, Republic of Korea

³Department of Life Science, Chung-Ang University, Seoul 156-756, Republic of Korea

(Received Feb 23 2015 / Revised Mar 13, 2015 / Accepted Mar 13, 2015)

Leuconostoc lactis EFEL005 (KACC 91922) isolated from kimchi showed promising probiotic attributes; resistance against acid and bile salts, absence of transferable genes for antibiotic resistance, broad utilization of prebiotics, and no hemolytic activity. To expand our understanding of the species, we generated a draft genome sequence of the strain and analyzed its genomic features related to the aforementioned probiotic properties. Genome assembly resulted in 35 contigs, and the draft genome has 1,688,202 base pairs (bp) with a G+C content of 43.43%, containing 1,644 protein-coding genes and 50 RNA genes. The average nucleotide identity analysis showed high homology ($\geq 96\%$) to the type strain *L. lactis* KCTC3528, but low homology ($\leq 95\%$) to *L. lactis* KCTC3773 (formerly *L. argentinum*). Genomic analysis revealed the presence of various genes for sucrose metabolism (glucansucrases, invertases, sucrose phosphorylases, and mannitol dehydrogenase), acid tolerance (F1F0 ATPases, cation transport ATPase, branched-chain amino acid permease, and lysine decarboxylase), vancomycin response regulator, and antibacterial peptide (Lactacin F). No gene for production of biogenic amines (histamine and tyramine) was found. This report will facilitate the understanding of probiotic properties of this strain as a starter for fermented foods.

Keywords: lactic acid bacteria, *Leuconostoc lactis*, probiotics, kimchi, draft genome

Introduction

Probiotics are defined as “live microorganisms that, when administered in adequate amounts, confer a health benefit on the host” (FAO/WHO), and currently several species of Lactobacilli, Lactococci, and Bifidobacteria are approved as probiotics by the Korean Ministry of Food and Drug Safety

(<http://www.foodnara.go.kr/hfoodi/industry/>). However, no species of *Leuconostoc* are regarded as probiotics, even though *L. mesenteroides* and *L. citreum* are important starter microbes used in several industrial and food fermentation processes, such as the production of cheese, butter, kefir, sourdough, and kimchi (De Bruyne *et al.*, 2007). The disqualification of this genus as probiotics is mainly due to the low colonization probability caused by the absence of acid and bile salt resistance. A phylogenetic analysis of the *Leuconostoc* genus based on 16S rRNA genes revealed that *L. lactis* is located between *L. mesenteroides*, and *L. citreum*, showing its relevance to the predominance species in kimchi (Hemme and Foucaud-Scheunemann, 2004)

Recently, we isolated a novel probiotic candidate, *L. lactis* EFEL005, from kimchi that has resistance to acid and bile salts (Noh, 2015). The strain quickly metabolizes diverse prebiotic substrates such as fructooligosaccharides, isomaltoligosaccharides, and xylooligosaccharides. These properties suggest that the strain can survive to reach the human large intestine and propagate to high density via the consumption of prebiotics ingested with meals. In addition, this strain is safe for use as a food additive because it shows no hemolytic activity against human red blood cells and lacks transferable antibiotic-resistance genes. Despite its potential, *L. lactis* was poorly characterized at the genome level. Therefore, in this study, we generated a draft genome sequence of the strain and analyzed its genomic features related to probiotic properties.

Materials and Methods

Bacterial culture and DNA extraction

L. lactis EFEL005 was isolated from kimchi in Cheongju city, Korea (36°38'N, 127°29'E). This strain was routinely grown in Lactobacilli MRS (Difco) broth or agar medium at 30°C. The culture was maintained in 50% glycerol solution at -70°C until needed. Dextran production was tested by growth on PES (phenylethyl alcohol containing 2% sucrose) agar. *L. lactis* EFEL005 was cultured in MRS medium at 30°C, and genomic DNA was extracted using a QIAamp DNA Mini Kit (Qiagen), following the standard protocol recommended by the manufacturer.

Phylogenetic analysis

The 16S rRNA gene of EFEL005 was amplified using the universal primers 27F and 1492R (Lane, 1991), and was sequenced by BIOFACT Co. The sequences of the reference strains used for phylogenetic analysis were obtained from

*For correspondence. E-mail: namsoo@chnugbuk.ac.kr; Tel.: +82-43-261-2567; Fax: +82-43-271-4412

Table 1. Comparison of the major genomic features of *L. lactis* EFEL005 and other *Leuconostoc* strains

Genome	Length (bp)	G+C Content (%)	Predicted ORFs	rRNA genes	tRNA genes	Isolated from	SignalP (%)	References
<i>L. lactis</i> KCTC 3528	2,011,205	42.64	2,727	3	46	Milk	2.78	Unpublished
<i>L. lactis</i> EFEL005	1,688,202	43.43	1,644	3	50	Kimchi	4.87	In this study
<i>L. argentinum</i> KCTC 3773	1,720,683	42.89	1,759	3	48	Milk	3.92	Nam <i>et al.</i> (2010)
<i>L. carnosum</i> JB16	1,773,468	37.13	1,691	12	66	Kimchi	3.96	Jung <i>et al.</i> (2012a)
<i>L. citreum</i> KM20	1,896,614	38.87	1,820	12	70	Kimchi	4.23	Kim <i>et al.</i> (2008)
<i>L. gasicomitatum</i> LMG 18811	1,954,080	36.66	1,912	12	67	Meat	3.77	Johansson <i>et al.</i> (2011)
<i>L. gelidum</i> JB7	1,893,499	36.68	1,796	12	67	Kimchi	3.56	Jung <i>et al.</i> (2012b)
<i>L. kimchii</i> IMSNU 11154	1,877,273	37.90	1,855	12	68	Kimchi	4.37	Oh <i>et al.</i> (2010)
<i>Leuconostoc</i> sp. C2	2,101,787	37.91	2,129	12	68	Kimchi	4.18	Lee <i>et al.</i> (2011)
<i>L. mesenteroides</i> ATCC 8293	2,075,763	37.67	2,005	12	71	Olives	4.94	Makarova <i>et al.</i> (2006)
<i>L. mesenteroides</i> J18	2,016,426	37.65	1,937	12	71	Kimchi	4.49	Jung <i>et al.</i> (2012c)

the EzTaxon-e (<http://eztaxon-e.ezbiocloud.net>; Kim *et al.*, 2012) or GenBank (Benson *et al.*, 2014) databases. Sequence similarities were calculated using the EzTaxon-e server. Multiple alignment with closely related sequences was performed using CLUSTAL W (Thompson *et al.*, 1997). A phylogenetic tree was generated using the neighbor-joining method (Saitou and Nei, 1987) using MEGA 5.2 software (Tamura *et al.*, 2011). A bootstrap analysis using 1,000 replicates was performed to assess the confidence limits of phylogenetic trees (Felsenstein, 1985).

Genome sequencing and assembly

Genomic DNA of *L. lactis* EFEL005 was sequenced by Chun-Lab using the Illumina MiSeq platform with a 300-bp paired-end library. The Illumina reads were assembled using CLC Genomics Workbench 6.0 (CLCbio). The initial assembly was converted for the CLC Genomics Workbench by constructing artificial reads from the consensus to collect the read pairs in the Illumina paired-end library. CodonCode Aligner 3.7.1 (CodonCode Corp.) was used for sequence assembly

and quality assessment in the subsequent finishing process. A total of 7,066,614 sequencing reads from the Illumina platform provided 775.85× coverage of the genome.

Genome annotation

The coding sequences (CDSs) were predicted by Glimmer 3.02 (Delcher *et al.*, 2007). The tRNAs were identified by tRNA-Scan-SE (Lowe and Eddy, 1997), and rRNAs were found using HMMER with EzTaxon-e rRNA profiles (Kim *et al.*, 2012). For functional annotation, the predicted CDSs were compared to catalytic families (catFam) and NCBI Clusters of Orthologous Groups (COGs) by rpsBLAST and NCBI reference sequences (RefSeq), and to SEED databases by BLASTP (Tatusov *et al.*, 2000; Overbeek *et al.*, 2005; Pruitt *et al.*, 2009; Yu *et al.*, 2009).

Comparison genome analysis

Eleven genome sequences (Table 1) of strains that belong to *Leuconostoc* species were obtained from the EzGenome

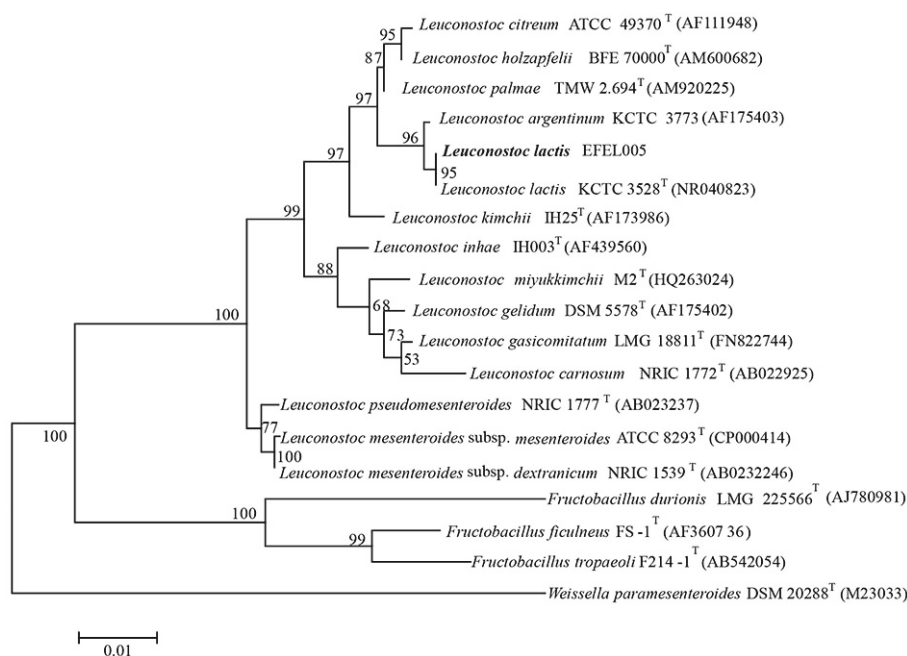


Fig. 1. Neighbor-joining tree generated using MEGA 6.0 based on 16S rRNA gene sequences, showing the phylogenetic relationships of *L. lactis* EFEL005 and related taxa. Bootstrap values are shown as percentages of 1,000 replicates when these values are greater than 50%. *Weissella paramesenteroides* DSM 20288^T (M23033) was used as an outgroup. Bar, 0.01 changes per nucleotide position.

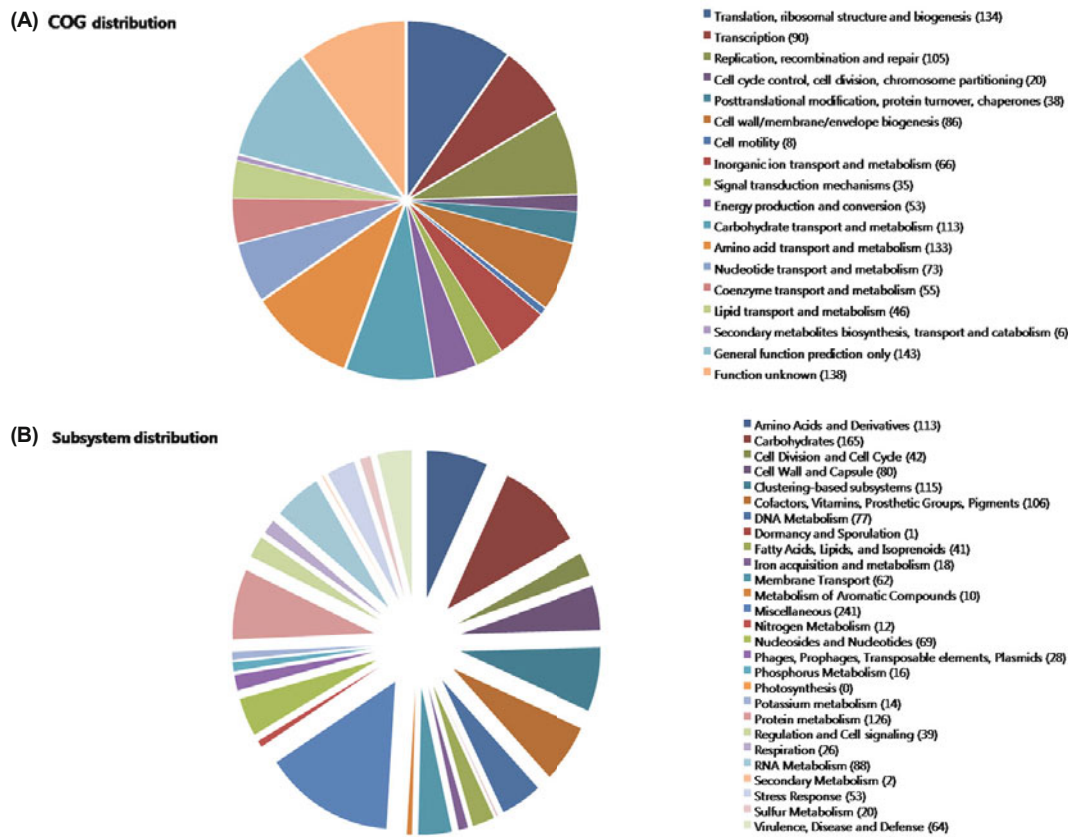


Fig. 2. Statistics of annotated genes for *L. lactis* EFEL005 based on the COG (A) and SEED (B) databases.

database (<http://ezgenome.ezbiocloud.net>), and used to calculate average nucleotide identity (ANI) values (Goris *et al.*, 2007) with strain EFEL005. ANI is a similarity measure between two genome sequences that may be used to replace

DNA-DNA hybridization. For ANI calculation, the query genome was cut into small fragments (1,020 bp), and high-scoring pairs between two genome sequences were selected using the BLAST algorithm (Altschul *et al.*, 1997).

Table 2. COG (cluster of orthologous groups of proteins) distribution of ORFs (open reading frames) found in a draft genome of *L. lactis* EFEL005

Category	Group	Function	Number of ORF
Information storage and processing	J	Translation, ribosomal structure and biogenesis	134
	K	Transcription	90
	L	DNA replication, recombination and repair	105
Cellular processes	D	Cell division and chromosome partitioning	20
	O	Posttranslational modification, protein turnover, chaperones	38
	M	Cell envelope, biogenesis, outer membrane	86
	N	Cell motility and secretion	8
	P	Inorganic ion transport and metabolism	66
	T	Signal transduction mechanisms	35
Cell metabolism	C	Energy production and conversion	53
	G	Carbohydrate transport and metabolism	113
	E	Amino acid transport and metabolism	133
	F	Nucleoside transport and metabolism	73
	H	Coenzyme metabolism	55
	I	Lipid metabolism	46
Others	Q	Secondary metabolites biosynthesis, transport and catabolism	6
	R	General function prediction only	143
Total	S	Function unknown	138
			1342

Table 3. Average nucleotide identity (ANI) values among 11 *Leuconostoc* strains

Strains	1	2	3	4	5	6	7	8	9	10	11
1. <i>L. argentinum</i> KCTC 3773	100.00	96.91	94.38	74.13	74.92	73.89	73.84	74.20	73.83	72.91	72.85
2. <i>L. lactis</i> KCTC 3528	96.07	100.00	96.10	74.13	74.08	73.55	73.48	74.26	73.51	73.03	72.87
3. <i>L. lactis</i> EFEL005	94.52	97.34	100.00	73.49	73.96	73.73	73.62	73.85	73.51	72.19	71.88
4. <i>L. carnosum</i> JB16 ^a	74.19	74.09	73.75	100.00	74.64	75.95	75.75	75.51	74.80	73.96	74.48
5. <i>L. citreum</i> KM20 ^a	74.92	75.29	74.17	74.48	100.00	74.42	74.28	74.37	73.88	73.20	73.28
6. <i>L. gasicomitatum</i> LMG 18811 ^a	73.97	74.35	73.83	76.02	74.32	100.00	94.75	77.36	77.27	73.19	73.13
7. <i>L. gelidum</i> JB7 ^a	73.95	74.20	73.93	75.77	74.47	94.88	100.00	77.47	77.34	73.21	73.21
8. <i>L. kimchii</i> IMSNU 11154 ^a	74.15	75.31	74.23	75.52	74.58	77.45	77.47	100.00	99.31	74.32	74.33
9. <i>Leuconostoc</i> sp. C2 ^a	74.11	74.23	73.80	74.70	73.92	77.24	77.26	99.51	100.00	73.03	72.97
10. <i>L. mesenteroides</i> ATCC 8293 ^a	72.50	73.09	72.21	73.84	72.98	73.27	73.30	74.44	72.87	100.00	98.98
11. <i>L. mesenteroides</i> J18 ^a	72.57	74.01	72.01	74.43	73.20	73.05	73.13	74.33	72.85	99.12	100.00

^a Complete sequences

ANI values >95%, a de facto cut off value for the identical species, are highlighted in bold.

Results and Discussion

Microbial features and classification

L. lactis EFEL005 is Gram-positive, non-motile, and sphere-shaped, 1.0–1.5 µm in length, and 0.4–0.5 µm in diameter (Garvie et al., 1986). It produced dextran within 1–2 days of growth on PES agar at 30°C. LAB gene sequences showing more than 99.7% sequence identity to the 16S rRNA of *L. lactis* KCTC 3528^T were selected and aligned for phylogenetic analysis using the ClustalW algorithm (Fig. 1).

General genome properties

The genome consists of a chromosome with a length of 1,688,202 bp and a G+C content of 43.43%. Of the 1,694 predicted genes, 1,644 are protein-coding genes and 50 are RNA genes. The annotation of each CDS was made by homology search against the cluster of orthologous groups of proteins (COG) and SEED databases (Disz et al., 2010). Results of the genome annotation are shown in Fig. 2. In the COG distribution, R (general function prediction only; 143 open reading frames [ORFs]), S (function unknown; 138 ORFs), and E (amino acid transport and metabolism; 133 ORFs) were abundant categories. Genes responsible for miscellaneous (241 ORFs), carbohydrates (165 ORFs), clustering-based subsystems (115 ORFs), and amino acid metabolism (113 ORFs) were abundant among the SEED subsystem categories. Three hundred twenty-nine genes were assigned to be involved in information storage and processing, and another 253 genes were identified to participate in cellular processes. The most abundant 479 genes encoded the proteins for cell metabolism, but another 281 genes were poorly characterized (Table 2).

DNA relatedness was previously used as an indicator to identify a new prokaryote species (Gardan et al., 1999); however, in this era of genomics, ANI between a given pair of genomes has become the preferred option (Goris et al., 2007). The ANI value between EFEL005 and KCTC 3528^T (a type strain) was greater than 96% (Table 3), indicating that the two strains are the same species; the proposed cut-off for species boundary is 95–96% (Richter and Rosselló-Móra, 2009).

Genomic features related to probiotic properties

Genomic analysis revealed the presence of various sucrose-catabolizing genes, such as Llac_01250, 07440, and 07470 for glucansucrases; Llac_12740, 15260, and 15930 for invertases; Llac_02250 for sucrose phosphorylases; and Llac_10130 for mannitol production. The results indicate that *L. lactis* EFEL005 has been well adapted in the plant-derived environment. Genes responsible for acid tolerance were also found: Llac_11300, 11310, 11320, 11330, 11340, 11350, 11360, and 11370 for F₁F₀ ATPases (Ventura et al., 2004); Llac_03800, 06210, 11760, and 16120 for cation transport ATPase (Kullen and Klaenhammer, 1999); Llac_01150 for branched-chain amino acid permease (Den Hengst et al., 2005); and Llac_06370 for lysine decarboxylase (Park et al., 1996). In addition, vancomycin response regulator gene (VanR) was found (Llac_14040). No genes for production of biogenic amines (histamine and tyramine) were found (Ammor and Mayo, 2007). The antibacterial peptide (Lactacin F) gene was Llac_10620 (Table 4).

In conclusion, we determined a draft genome sequence of *L. lactis* EFEL005, and examined its genomic features as a lactic acid bacterium. Our results will help to understand

Table 4. Gene list for probiotic properties in *L. lactis* EFEL005

Deduced function	Gene accession number
F ₁ F ₀ ATPases	Llac_11300
	Llac_11310
	Llac_11320
	Llac_11330
	Llac_11340
	Llac_11350
	Llac_11360
Cation transport ATPase	Llac_03800
	Llac_06210
	Llac_11760
	Llac_16120
Branched-chain amino acid permeases	Llac_01150
Arginine/lysine/ornithine decarboxylases	Llac_06370
Vancomycin response regulator VanR	Llac_14040
Antibacterial peptide (Lactacin F)	Llac_10620

the probiotic properties of this strain, and will contribute to use of the strain as a probiotic after additional *in vivo* studies.

Availability of supporting data

The genome sequence of *L. lactis* EFEL005 was deposited at DDBJ/EMBL/NCBI under the accession number JMEA-00000000. The version described in this paper is the first version JMEA01000000. The genome project for this strain is listed in the Genome On Line Database (GOLD) (Pagani *et al.*, 2012) under project Gi0073879.

Acknowledgements

This work was supported by the grants from the Korean Rural Development Administration (RDA) (PJ90715303) and Korea Institute of Planning and Evaluation for Technology in Food, Agriculture, Forestry and Fisheries (iPET,113034-3).

References

- Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402.
- Ammor, M.S. and Mayo, B. 2007. Selection criteria for lactic acid bacteria to be used as functional starter cultures in dry sausage production: An update. *Meat Sci.* **76**, 138–146.
- Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Sayers, E.W. 2014. GenBank. *Nucleic Acids Res.* **42**, 32–37.
- De Bruyne, K., Schillinger, U., Caroline, L., Boehringer, B., Cleenwerck, I., Vancanneyt, M., De Vuyst, L., Franz, C.M.A.P., and Vandamme, P. 2007. *Leuconostoc holzapfelii* sp. nov., isolated from Ethiopian coffee fermentation and assessment of sequence analysis of housekeeping genes from delineation of *Leuconostoc* species. *Int. J. Syst. Evol. Microbiol.* **57**, 2952–2959.
- Delcher, A.L., Bratke, K.A., Powers, E.C., and Salzberg, S.L. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* **23**, 673–679.
- Den Hengst, C.D., Van Hijum, S.A.F.T., Geurts, J.M.W., Nauta, A., Kok, J., and Kuipers, O.P. 2005. The *Lactococcus lactis* CodY regulon: Identification of a conserved cis-regulatory element. *J. Biol. Chem.* **280**, 34332–34342.
- Disz, T., Akhter, S., Cuevas, D., Olson, R., Overbeek, R., Vonstein, V., Stevens, R., and Edwards, R.A. 2010. Accessing the SEED genome databases via Web services API: tools for programmers. *BMC Bioinformatics* **11**, 319.
- Felsenstein, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **39**, 783–791.
- Gardan, L., Shafik, H., Belouin, S., Broch, R., Grimont, F., and Grimont, P.A.D. 1999. DNA relatedness among the pathovars of *Pseudomonas syringae* and description of *Pseudomonas tremae* sp. nov. and *Pseudomonas cannabina* sp. nov. (ex Sutic and Dowson 1959). *Int. J. Syst. Evol. Microbiol.* **49**, 469–478.
- Garvie, E.I. 1986. Genus *Leuconostoc* van Tieghem 1878, pp. 1071–1075. In Sneath, N.S., Mair, H.A., Sharpe, M.E., and Holt, J.G. (eds.), *Bergey's Manual of Systematic Bacteriology*, 8th ed. The Williams & Wilkins Co. Baltimore, Maryland, USA.
- Goris, J., Konstantinidis, K.T., Klappenbach, J.A., Coenye, T., Vandamme, P., and Tiedje, J.M. 2007. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* **57**, 81–91.
- Hemme, D. and Foucaud-Scheunemann, C. 2004. *Leuconostoc*, characteristics, use in dairy technology and prospects in functional foods. *Int. Dairy J.* **14**, 467–494.
- Johansson, P., Paulin, L., Säde, E., Salovuori, N., Alatalo, E.R., Björkroth, K.J., and Auvinen, P. 2011. Genome sequence of a food spoilage lactic acid bacterium, *Leuconostoc gasicomitatum* LMG 18811^T, in association with specific spoilage reactions. *Appl. Environ. Microbiol.* **77**, 4344–4351.
- Jung, J.Y., Lee, S.H., and Jeon, C.O. 2012a. Complete genome sequence of *Leuconostoc carnosum* strain JB16, isolated from kimchi. *J. Bacteriol.* **194**, 6672–6673.
- Jung, J.Y., Lee, S.H., and Jeon, C.O. 2012b. Complete genome sequence of *Leuconostoc gelidum* strain JB7, isolated from kimchi. *J. Bacteriol.* **194**, 6665.
- Jung, J.Y., Lee, S.H., Lee, S.H., and Jeon, C.O. 2012c. Complete genome sequence of *Leuconostoc mesenteroides* subsp. *mesenteroides* strain J18, isolated from kimchi. *J. Bacteriol.* **194**, 730–731.
- Kim, J.F., Jeong, H., Lee, J.S., Choi, S.H., Ha, M., Hur, C.G., Kim, J.S., Lee, S., Park, H.S., Park, Y.H., *et al.* 2008. Complete genome sequence of *Leuconostoc citreum* KM20. *J. Bacteriol.* **190**, 3093–3094.
- Kim, O.S., Cho, Y.J., Lee, K., Yoon, S.H., Kim, M., Na, H., Park, S.C., Jeon, Y.S., Lee, J.H., Yi, H., *et al.* 2012. Introducing EzTaxon-e: A prokaryotic 16S rRNA gene sequence database with phylogenies that represent uncultured species. *Int. J. Syst. Evol. Microbiol.* **62**, 716–721.
- Kullen, M.J. and Klaenhammer, T.R. 1999. Identification of the pH-inducible, proton-translocating F1F0-ATPase (atpBEFHAGDC) operon of *Lactobacillus acidophilus* by differential display: Gene structure, cloning and characterization. *Mol. Microbiol.* **33**, 1152–1161.
- Lane, D.J. 1991. 16S/23S rRNA sequencing, pp. 115–175. In Stackebrandt, E. and Goodfellow, M.D. (eds.), *Nucleic Acid Techniques in Bacterial Systematics*. John Wiley and Sons, New York, N.Y., USA.
- Lee, S.H., Jung, J.Y., Lee, S.H., and Jeon, C.O. 2011. Complete genome sequence of *Leuconostoc kimchii* strain C2, isolated from kimchi. *J. Bacteriol.* **193**, 5548.
- Lowe, T.M. and Eddy, S.R. 1997. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**, 955–964.
- Makarova, K., Slesarev, A., Wolf, Y., Sorokin, A., Mirkin, B., Koonin, E., Pavlov, A., Pavlova, N., Karamychev, V., Polouchine, N., *et al.* 2006. Comparative genomics of the lactic acid bacteria. *Proc. Natl. Acad. Sci. USA* **103**, 15611–15616.
- Nam, S.H., Choi, S.H., Kang, A., Kim, D.W., Kim, R.N., Kim, A., and Park, H.S. 2010. Genome sequence of *Leuconostoc argentinum* KCTC 3773. *J. Bacteriol.* **192**, 6490–6491.
- Noh, S.J. 2015. Master's thesis. Chungbuk National University, Cheongju, Chungbuk, Korea.
- Oh, H.M., Cho, Y.J., Kim, B.K., Roe, J.H., Kang, S.O., Nahm, B.H., Jeong, G., Han, H.U., and Chun, J. 2010. Complete genome sequence analysis of *Leuconostoc kimchii* IMSNU 11154. *J. Bacteriol.* **192**, 3844–3845.
- Overbeek, R., Begley, T., Butler, R.M., Choudhuri, J.V., Chuang, H.Y., Cohoon, M., de Crécy-Lagard, V., Diaz, N., Disz, T., Edwards, R., *et al.* 2005. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res.* **33**, 5691–5702.
- Pagani, I., Liolios, K., Jansson, J., Chen, I.M.A., Smirnova, T., Nosrat, B., Markowitz, V.M., and Kyrpides, N.C. 2012. The Genomes OnLine Database (GOLD) v.4: Status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res.* **40**, 571–579.
- Park, Y.K., Bearson, B., Bang, S.H., Bang, I.S., and Foster, J.W. 1996. Internal pH crisis, lysine decarboxylase and the acid tolerance response of *Salmonella typhimurium*. *Mol. Microbiol.* **20**, 605–611.
- Pruitt, K.D., Tatusova, T., Klimke, W., and Maglott, D.R. 2009.

- NCBI reference sequences: Current status, policy and new initiatives. *Nucleic Acids Res.* **37**, 32–36.
- Richter, M. and Rosselló-Móra, R.** 2009. Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. USA* **106**, 19126–19131.
- Saitou, N. and Nei, M.** 1987. The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425.
- Tamura, K., Dudley, J., Nei, M., and Kumar, S.** 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731–2739.
- Tatusov, R.L., Galperin, M.Y., Natale, D.A., and Koonin, E.V.** 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* **28**, 33–36.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G.** 1997. The CLUSTAL X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**, 4876–4882.
- Ventura, M., Canchaya, C., Zink, R., Fitzgerald, G.F., and Van Sinderen, D.** 2004. Characterization of the groEL and groES loci in *Bifidobacterium breve* UCC 2003: Genetic, transcriptional, and phylogenetic analyses. *Appl. Environ. Microbiol.* **70**, 6197–6209.
- Yu, C., Zavaljevski, N., Desai, V., and Reifman, J.** 2009. Genome-wide enzyme annotation with precision control: Catalytic families (CatFam) databases. *Proteins Struct. Funct. Bioinforma.* **74**, 449–460.