



Genomic Analysis of *Bacillus licheniformis* CBA7126 Isolated from a Human Fecal Sample

Changsu Lee¹, Joon Yong Kim¹, Hye Seon Song¹, Yeon Bee Kim¹, Yoon-E Choi², Changmann Yoon², Young-Do Nam^{3,4*} and Seong Woon Roh^{1*}

¹ Microbiology and Functionality Research Group, World Institute of Kimchi, Gwangju, South Korea, ² Division of Environmental Science & Ecological Engineering, Korea University, Seoul, South Korea, ³ Research Group of Gut Microbiome, Korea Food Research Institute, Sungnam, South Korea, ⁴ Department of Food Biotechnology, University of Science and Technology, Daejeon, South Korea

Keywords: *Bacillus licheniformis*, genome sequence, human fecal sample, stress response genes, multilocus sequence typing

INTRODUCTION

Bacillus licheniformis is a Gram-positive, endospore-forming, saprophytic organism that occurs in plant and soil (Veith et al., 2004). A taxonomical approach shows that it is closely related to *Bacillus subtilis* (Lapidus et al., 2002; Xu and Côte, 2003; Rey et al., 2004). Generally, most bacilli are predominantly aerobic; however, *B. licheniformis* is a facultative anaerobe compared to other bacilli in ecological niches (Alexander, 1977). The commercial utility of the extracellular products of *B. licheniformis* makes this microorganism an economically interesting species (Kovács et al., 2009). For example, *B. licheniformis* is used industrially for manufacturing biochemicals, enzymes, antibiotics, and aminopeptidase. Several proteases such as α -amylase, penicillinase, pentosanase, cycloglucosyltransferase, β -mannanase, and certain pectinolytic enzymes are synthesized industrially using *B. licheniformis* (Rodríguez-Absi and Prescott, 1978; Rey et al., 2004). The proteases are used in the detergent industry and the amylases are utilized for starch hydrolysis, desizing of textiles, and sizing of paper (Erickson, 1976). In addition, certain strains are utilized to produce peptide antibiotics, specialty chemicals, and poly- γ -glutamic acid (Nierman and Maglott, 1989; Rey et al., 2004).

The annotated genome sequence of *B. licheniformis* has been previously analyzed to assess the biotechnological importance of the organism (Veith et al., 2004). Since the first sequencing, the genomes of specific *B. licheniformis* strains have been sequenced to completely realize its industrial potential. In this study, genome sequencing of *B. licheniformis* CBA7126 isolated from a human fecal sample was performed to understand bacterial specificity. The genome sequence of CBA7126 revealed features such as stress response genes, antibiotic-resistance genes, and genes for resistance to toxic compounds, which are of considerable biotechnological value.

MATERIALS AND METHODS

Bacterial Isolation, Culture Conditions, and DNA Extraction

B. licheniformis CBA7126 was isolated from the feces of a 74-year-old man in Geochang-gun, South Korea and was cultured under anaerobic conditions in Gifu Anaerobic Medium (GAM) (containing per liter of deionized distilled water: 10 g peptone, 3 g soytone, 10 g proteose peptone, 13.5 g bovine serum albumin, 5 g yeast extract, 2.2 g beef extract, 2.5 g monopotassium phosphate, 1.2 g liver extract, 3 g sodium chloride, 0.3 g L-cystein, 0.3 g sodium thioglycollate, 3 g dextrose, 5 g soluble starch) at 37°C for 48 h. Genomic DNA of strain CBA7126 was extracted using the QIAamp DNA extraction kit (Qiagen, USA) and QuickGene DNA tissue kit S (Kurabo, Japan),

OPEN ACCESS

Edited by:

Annalisa Bruno,
Università degli Studi "G. d'Annunzio"
Chieti - Pescara, Italy

Reviewed by:

Vasvi Chaudhry,
Institute of Microbial Technology
(CSIR), India
Georgios Paschos,
University of Pennsylvania,
United States

*Correspondence:

Young-Do Nam
youngdo98@kfri.re.kr
Seong Woon Roh
seong18@gmail.com

Specialty section:

This article was submitted to
Inflammation Pharmacology,
a section of the journal
Frontiers in Pharmacology

Received: 01 August 2017

Accepted: 25 September 2017

Published: 13 October 2017

Citation:

Lee C, Kim JY, Song HS, Kim YB,
Choi Y-E, Yoon C, Nam Y-D and
Roh SW (2017) Genomic Analysis of
Bacillus licheniformis CBA7126
Isolated from a Human Fecal Sample.
Front. Pharmacol. 8:724.
doi: 10.3389/fphar.2017.00724

and purified using the MG genomic DNA purification kit (Doctor Protein, Korea) according to the manufacturer's instructions. The purity and concentration of the extracted genomic DNA were measured using the Nanodrop spectrophotometer (NanoDrop Technologies, UK).

Genome Sequencing, Assembly, and Annotation

The genome of *B. licheniformis* CBA7126 was sequenced using a 20-kb SMRTbell library and PacBio RS II system (Pacific Biosciences, USA), and *de novo* assembly was performed using the HGAP2 protocol in PacBio SMRT Analysis version 2.3.0. rRNAs and tRNAs were analyzed using RNAmmer 1.2 (Lagesen et al., 2007) and tRNAscan-SE 1.21 (Lowe and Eddy, 1997), respectively. The potential coding regions and functional genes were predicted via a combination of Glimmer 3.02 (Delcher et al., 1999), COG database (Tatusov et al., 2003), the Rapid Annotation Search Tool (RAST) (Aziz et al., 2008), and the National Center for Biotechnology Information (NCBI) prokaryotic genome annotation pipeline (PGAP) 4.1 (Tatusova et al., 2016). Prophages in the genome were identified using the PHAge Search Tool (PHAST) (Zhou et al., 2011). In addition, pathogenicity of strain CBA7126 was predicted using PathogenFinder 1.1 (Cosentino et al., 2013). Carbohydrate-active enzymes were annotated using dbCAN (Yin et al., 2012).

Comparative Genomic Analysis

To identify the unique features of strain CBA7126, the genomes of *B. licheniformis* and *Bacillus* sp. strains (*B. licheniformis* B4164, *B. licheniformis* VTM3R78, *B. licheniformis* V30, *B. licheniformis* B4124, and *Bacillus* sp. H15-1) were selected for comparative genomic analysis using the NCBI genome database (<http://www.ncbi.nlm.nih.gov/genome/>). For calculation of overall genome relatedness, average nucleotide identity (ANI), and orthologous average nucleotide identity (OrthoANI) analysis of *B. licheniformis* CBA7126 was performed on sequences of related species using the ANI calculator (<http://enve-omics.ce.gatech.edu/ani/>) and orthologous average nucleotide identity tool (OAT) of ChunLab (Lee et al., 2016). The genome structure of strain CBA7126 was compared to those of *B. licheniformis* B4164 (LQYQ00000000.1), *B. licheniformis* VTM3R78 (FOFE00000000.1), *B. licheniformis* V30 (LQRR00000000.1), *Bacillus* sp. H15-1 (CP018249.1), and *B. licheniformis* B4124 (LKPQ00000000.1) having symmetric identity >97% with strain CBA7126, using the alignment program MAUVE (Darling et al., 2004). Pan-genome Orthologous Groups (POGs) were analyzed using BIOiPLUG Comparative Genomics Database (<https://www.bioiplug.com/>). Venn diagram was constructed based on the number of POGs of strain CBA7126 and the related strains. Clustered regularly interspaced short palindromic repeats (CRISPR) was analyzed using CRISPRfinder (Grissa et al., 2007).

Multilocus Sequence Typing (MLST)

Multilocus sequence typing (MLST) analysis based on internal sequences of *adk*, *ccpA*, *recF*, *rpoB*, *spo0A*, and *sucC* genes was performed (Larsen et al., 2012; Madslie et al., 2012). The MLST

sequence type of strain CBA7126 was determined using the MLST 1.8 database (<https://cge.cbs.dtu.dk/services/MLST/>) of *B. licheniformis* (Larsen et al., 2012).

Ethics Approval

The study protocol was approved by the institutional review board of the Theragen ETEX Bio Institute (700062-20160804-JR-005-02).

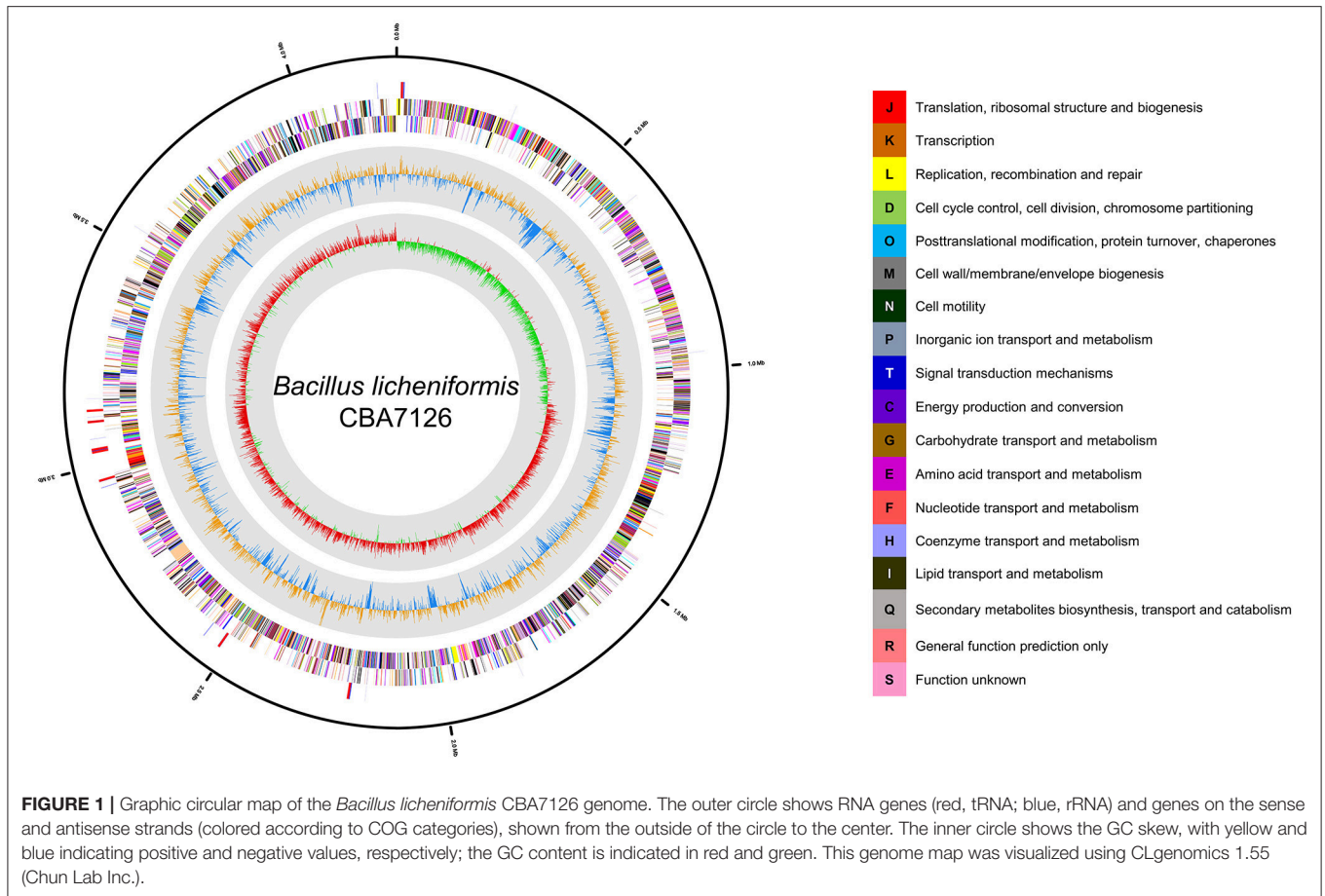
RESULTS

General Genomic Features of *B. licheniformis* CBA7126

The genome of *B. licheniformis* CBA7126 was 4,216,391 bp long with a G + C content of 46.24 mol% (Table 1). The genome is predicted to contain two contigs of 4,209,959 and 6,972 bp. Strain CBA7126 genome contained 4,276 coding sequences, 24 rRNA genes (8 of the 16S-5S-23S RNA gene operon), and 81 tRNA genes (Figure 1). For functional classification, the genome of strain CBA7126 was analyzed using the Cluster of Orthologous Groups (COG) database (<http://www.ncbi.nlm.nih.gov/COG/>), and 3,743 genes were annotated. The annotated genes belonged to the following categories: function unknown (S; 884 genes), general function prediction only (R; 344), transcription (K; 319 genes), carbohydrate transport, and metabolism (G; 316 genes), amino acid transport and metabolism (E; 298 genes), inorganic ion transport and metabolism (P; 219 genes), energy production and conversion (C; 180 genes), replication, recombination, and repair (L; 140 genes), and secondary metabolite biosynthesis, transport, and catabolism (Q; 62 genes) (Supplementary Table 1). In addition, SEED viewer version 2.0 revealed that >9% of the major categories contained genes required for metabolism of "carbohydrates" (610 genes), "amino acids and derivatives" (457 genes), and "cofactors, vitamins, prosthetic group, pigments" (280 genes) (Supplementary Figure 1). A total of 193 CAZyme-encoding genes were annotated using dbCAN, including five for auxiliary activities (AAs), 39 for carbohydrate-binding modules (CBMs), 36 for carbohydrate esterases (CEs), 68 for glycoside hydrolases (GHs), 39 for glycosyl transferases (GTs), and 6 for polysaccharide lyases (PLs).

Comparative Genomic Data

Analysis of the orthoANI values among *Bacillus* genome sequences with symmetric identity of >97% revealed that *B. licheniformis* CBA7126 has higher than 99% genome sequence similarity with other species. The genome of strain CBA7126 was closest to that of *B. licheniformis* VTM3R78 (99.99% orthoANI), followed by *B. licheniformis* B4164 (99.98%), *Bacillus* sp. H15-1 (99.85%), *B. licheniformis* B4124 (99.81%), and *B. licheniformis* V30 (99.80%) (Supplementary Figure 2). Similar results were also obtained using ANI. Based on the results of Lee et al. (2016), similarity values >95–96% indicated that two strains belong to the same species. Therefore, strain CBA7126 was confirmed to be a species of *B. licheniformis*. The genome of strain CBA7126 was aligned with more than 97% symmetric identity with those of strains *B. licheniformis* B4164, *B. licheniformis* VTM3R78, *B. licheniformis* V30, *Bacillus* sp. H15-1, and *B. licheniformis* B4124



using MAUVE. The genomic representations of the other strains were rearranged based on the structure of strain CBA7126. Gene order comparison was established for seven regions with Local Collinear Blocks (LCBs). The structure of strain CBA7126 was similar to that of *B. licheniformis* B4124 and *B. licheniformis* V30 (Supplementary Figure 3). Comparison of strain CBA7126 genomic structure with that of *Bacillus* sp. H15-1 showed that two major regions were in opposite direction. Analysis based on the POG of strain CBA7126 and the closely related strains identified 4,108 shared genes and 137 unique genes (Supplementary Figure 4). Strain CBA7126 possessed 19 genes among the unique genes: 1 poly (glycerol-phosphate) alpha-glucosyltransferase, 1 thymidylate synthase (FAD), 2 prophage-derived protein, and 15 hypothetical proteins. The three genes among the 19 unique genes of strain CBA7126 were classified to one carbon pool by folate, pyrimidine metabolism, and metabolic pathways, based on KEGG analysis. In addition, CRISPR analysis indicated that strain CBA7126 did not harbor any known CRISPRs.

Phage and Pathogenesis-Related Genes

PHAST analysis was performed for identifying prophage contamination in the genome sequence of strain CBA7126. Contig 1 contained three intact and two incomplete prophages, whereas contig 2 contained only one incomplete prophage (Supplementary Figure 5). Intact regions of prophages

were located between positions 1,596,547–1,623,555, 1,775,723–1,820,161, and 3,429,284–3,483,201 bp, respectively. Strain CBA7126 was identified to be a human pathogen with 0.81 probability in PathogenFinder 1.1. Analysis of pathogenesis-related genes showed that all the 238 analyzed genes encoded pathogenesis-associated proteins.

Multilocus Sequence Typing (MLST) Analysis

MLST analysis of strain CBA7126 was performed using six housekeeping genes (*adk*, *ccpA*, *recF*, *rpoB*, *spo0A*, and *sucC*). MLST analysis showed that strain CBA7126 belonged to sequence type 3 since this organism harbored *adk_2*, *ccpA_1*, *recF_1*, *rpoB_1*, *spo0A_1*, and *sucC_2* (Supplementary Table 2). Previously reported isolates of sequence type 3 are *B. licheniformis* NVH1023, F5520, CCUG41412, NVH1111, NVH1113, LMG17661, and M3.

Stress Response Genes and Resistance to Toxic Compounds

Comparison with NCBI PGAP 4.1 showed that the genome of strain CBA7126 harbors several stress response genes and various genes required for resistance to antibiotics and toxic compounds (Tatusova et al., 2016). The identified

TABLE 1 | Features of the *Bacillus licheniformis* CBA7126 genome.

	<i>B. licheniformis</i> CBA7126
Sequencing platform	PacBio RS II system
Assembler	PacBio SMRT Analysis 2.3.0
Assembly accession	GCA_001950175.1
Methods reads	90,824
Assembly size (bp)	4,216,931
Contig numbers	2
N50	4,209,959
L50	1
Genome coverage	319.23
DNA G + C content (mol%)	46.24
CDSs	4,276
rRNA number	24
tRNA number	81
Genes assigned to COGs	3,743
CRISPRs	0

stress tolerance genes encode general stress proteins (WP_003179040.1; WP_009329495.1; WP_011198337.1; WP_003186243.1), universal stress proteins (WP_011197701.1; WP_003178013.1), cold shock proteins (WP_003153604.1; WP_003179166.1), and the UV-damage repair protein UvrX (WP_003183238.1). These genes were closely associated with the survival of bacteria in the natural environment. The genes identified for resistance to toxic compounds encode monooxygenase (required for antibiotic resistance) (WP_017473926.1; WP_003181975.1), L-asparaginase (WP_003183042.1; WP_003183042.1; WP_061565867.1), the multidrug resistance protein NorM (WP_009328059.1), YkkD (WP_003180981.1), YkkC (WP_003180979.1), arginase (WP_009330115.1; WP_009330115.1; WP_003178878.1; WP_003178436.1), chemical damaging agent resistance protein C (WP_017474008.1; WP_003178723.1), toxic anion resistance

protein (WP_003178733.1), lantibiotic-related proteins (WP_003186355.1; WP_003186351.1; WP_003186379.1; WP_003186381.1), bacitracin, and various proteins of the ABC transporter family. Among the genes related to stress response, L-asparaginase, arginase, lantibiotic, and bacitracin are used for industrial application.

DATA ACCESS

The genome sequence of *B. licheniformis* CBA7126 has been deposited in DDBJ/ENA/GenBank under the accession numbers BDJJ01000001–BDJJ01000002.

AUTHOR CONTRIBUTIONS

SR and YN designed and coordinated all the experiments. HS performed cultivation, DNA extraction, and purification. CL, JK, HS, YK, YC, and CY performed the sequencing, genome assembly, gene prediction, gene annotation, and comparative genomic analysis. CL, YN, and SR wrote manuscript. All authors have read and approved the manuscript.

FUNDING

This research was supported by grants from the World Institute of Kimchi (KE1702-2), funded by the Ministry of Science and ICT, Korea Food Research Institute (E0170602-01), and Basic Science Research Program through the National Research Foundation of Korea (NRF) (2015R1D1A1A09061039), Republic of Korea.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fphar.2017.00724/full#supplementary-material>

REFERENCES

- Alexander, M. (1977). *Introduction to Soil Microbiology*. New York, NY: John Wiley & Sons.
- Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., et al. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 9:75. doi: 10.1186/1471-2164-9-75
- Cosentino, S., Larsen, M. V., Aarestrup, F. M., and Lund, O. (2013). Pathogenfinder-distinguishing friend from foe using bacterial whole genome sequence data. *PLoS ONE* 8:e77302. doi: 10.1371/annotation/b84e1af7-c127-45c3-be22-76abd977600f
- Darling, A. C., Mau, B., Blattner, F. R., and Perna, N. T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14, 1394–1403. doi: 10.1101/gr.2289704
- Delcher, A. L., Harmon, D., Kasif, S., White, O., and Salzberg, S. L. (1999). Improved microbial gene identification with GLIMMER. *Nucleic Acids Res.* 27, 4636–4641. doi: 10.1093/nar/27.23.4636
- Erickson, R. (1976). "Industrial applications of the Bacilli: a review and prospectus," in *Microbiology*, ed D. Schlesinger (Washington, DC: American Society for Microbiology), 406–419.
- Grissa, I., Vergnaud, G., and Pourcel, C. (2007). CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.* 35(Suppl. 2), W52–W57. doi: 10.1093/nar/gkm360
- Kovács, A. T., Smits, W. K., Mironczuk, A. M., and Kuipers, O. P. (2009). Ubiquitous late competence genes in *Bacillus* species indicate the presence of functional DNA uptake machineries. *Environ. Microbiol.* 11, 1911–1922. doi: 10.1111/j.1462-2920.2009.01937.x
- Lagesen, K., Hallin, P., Rødland, E. A., Stærfeldt, H.-H., Rognes, T., and Ussery, D. W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35, 3100–3108. doi: 10.1093/nar/gkm160
- Lapidus, A., Galleron, N., Andersen, J. T., Jørgensen, P. L., Ehrlich, S. D., and Sorokin, A. (2002). Co-linear scaffold of the *Bacillus licheniformis* and *Bacillus subtilis* genomes and its use to compare their competence genes. *FEMS Microbiol. Lett.* 209, 23–30. doi: 10.1111/j.1574-6968.2002.tb11104.x
- Larsen, M. V., Cosentino, S., Rasmussen, S., Friis, C., Hasman, H., Marvig, R. L., et al. (2012). Multilocus sequence typing of total genome sequenced bacteria. *J. Clin. Microbiol.* 50, 06094–06011. doi: 10.1128/JCM.06094-11
- Lee, I., Kim, Y. O., Park, S.-C., and Chun, J. (2016). OrthoANI: an improved algorithm and software for calculating average nucleotide identity. *Int. J. Syst. Evol. Microbiol.* 66, 1100–1103. doi: 10.1099/ijsem.0.000760

- 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. doi: 10.1093/nar/25.5.0955
- Madslie, E. H., Olsen, J. S., Granum, P. E., and Blatny, J. M. (2012). Genotyping of *B. licheniformis* based on a novel multi-locus sequence typing (MLST) scheme. *BMC Microbiol.* 12:230. doi: 10.1186/1471-2180-12-230
- Nierman, W., and Maglott, D. (1989). *American Type Culture Collection Catalogue of Bacteria and Bacteriophages*. Rockville, MD: American Type Culture Collection.
- Rey, M. W., Ramaiya, P., Nelson, B. A., Brody-Karpin, S. D., Zaretsky, E. J., Tang, M., et al. (2004). Complete genome sequence of the industrial bacterium *Bacillus licheniformis* and comparisons with closely related *Bacillus* species. *Genome Biol.* 5:r77. doi: 10.1186/gb-2004-5-10-r77
- Rodríguez-Absi, J., and Prescott, J. M. (1978). Isolation and properties of an aminopeptidase from *Bacillus licheniformis*. *Arch. Biochem. Biophys.* 186, 383–391. doi: 10.1016/0003-9861(78)90449-6
- Tatusov, R. L., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Kiryutin, B., Koonin, E. V., et al. (2003). The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4:41. doi: 10.1186/1471-2105-4-41
- Tatusova, T., DiCuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E. P., Zaslavsky, L., et al. (2016). NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* 44, 6614–6624. doi: 10.1093/nar/gkw569
- Veith, B., Herzberg, C., Steckel, S., Feesche, J., Maurer, K. H., Ehrenreich, P., et al. (2004). The complete genome sequence of *Bacillus licheniformis* DSM13, an organism with great industrial potential. *J. Mol. Microbiol. Biotechnol.* 7, 204–211. doi: 10.1159/000079829
- Xu, D., and Côte, J.-C. (2003). Phylogenetic relationships between *Bacillus* species and related genera inferred from comparison of 3' end 16S rDNA and 5' end 16S–23S ITS nucleotide sequences. *Int. J. Syst. Evol. Microbiol.* 53, 695–704. doi: 10.1099/ijs.0.02346-0
- Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., and Xu, Y. (2012). dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* 40, W445–W451. doi: 10.1093/nar/gks479
- Zhou, Y., Liang, Y., Lynch, K. H., Dennis, J. J., and Wishart, D. S. (2011). PHAST: a fast phage search tool. *Nucleic Acids Res.* 39(Suppl. 2), W347–W352. doi: 10.1093/nar/gkr485
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Copyright © 2017 Lee, Kim, Song, Kim, Choi, Yoon, Nam and Roh. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.