

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>Paenibacillus konkukensis</i> sp. nov. SK3146 as a potential probiotic strain
Running Title (within 10 words)	Complete genome sequence of <i>Paenibacillus konkukensis</i> sp. nov. SK3146
Author	Hae-In Jung ¹ , Sungkwon Park ² , Kai-Min Niu ³ , Sang-Won Lee ⁴ , Damini Kothari ¹ , Kwon Jung Yi ¹ , and Soo-Ki Kim ¹
Affiliation	1 Department of Animal Sciences and Technology, Konkuk University, Seoul 05029, Korea 2 Department of Food Science and Biotechnology, Sejong University, Seoul 05005, Korea 3 Institute of Biological Resources, Jiangxi Academy of Sciences, Nanchang 330029, China 4 College of Veterinary Medicine, Konkuk University, Seoul 05029, Korea
ORCID (for more information, please visit https://orcid.org)	Hae-In Jung (https://orcid.org/0000-0003-2643-6681) Sungkwon Park(https://orcid.org/0000-0002-7684-9719) Kai-Min Niu (https://orcid.org/0000-0002-9756-3517) Sang-Won Lee (https://orcid.org/0000-0003-1956-7245) Damini Kothari (https://orcid.org/0000-0003-3627-2377) Kwon Jung Yi (https://orcid.org/0000-0001-6002-0378) Soo-Ki Kim (https://orcid.org/0000-0003-3499-3330)
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.	Not applicable.
Acknowledgements	This research was supported by a grant from Agricultural Science and Technology Development Program (Project No. PJ010906), Rural Development Administration, Republic of Korea.
Availability of data and material	Upon reasonable request, the datasets of this study can be available from the corresponding author.
Authors' contributions Please specify the authors' role using this form.	Conceptualization: Jung HI, Lee SW, Park SK, Kim SK. Data curation: Jung HI. Formal analysis: Jung HI, Niu KM Methodology: Jung HI. Validation: Jung HI, Lee SW Investigation: Jung HI, Yi KJ Writing - original draft: Jung HI, Lee SW Writing - review & editing: Niu KM, Lee SW, Park SK, Kim SK.
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	Soo-Ki Kim
Email address – this is where your proofs will be sent	sookikim@konkuk.ac.kr

Secondary Email address	mebong7@hanmail.net
Address	Department of Animal Sciences and Technology, Konkuk University, Seoul 05029, Korea.
Cell phone number	+82-10-2965-3728
Office phone number	+82-2-450-3728
Fax number	+82-2-458-3728

ACCEPTED

Abstract

Paenibacillus konkukensis sp. nov., SK3146 is a novel strain isolated from a pig feed. Here, we present complete genome sequence of SK3146. The genome consists of a single circular genome measuring 7,968,964 bp in size with an average G+C content of 53.4%. Genomic annotation revealed that the strain encodes 151 proteins related to hydrolases (EC3), which was higher than those in *Bacillus subtilis* and *Escherichia coli*. Diverse kinds of hydrolases including galactosidase, glucosidase, cellulase, lipase, xylanase, and protease were found in the genome of SK3146, coupled with one bacteriocin encoding gene. The complete genome sequence of *P. konkukensis* SK3146 indicates the immense probiotic potential of the strain with nutrient digestibility and antimicrobial activity functions.

Keywords: *Paenibacillus*, Complete genome sequence, Exoenzyme, Feed additive

The well-being and health status of economic animals are more deteriorative due to the intensive farming practices. Feed-additives probiotics including lactic acid bacteria and *Bacillus* spp. are commonly used to modulate host health and improve performance in pig production [1, 2]. Typically, a culture-dependent isolation coupled with *in-vitro* characterization methods have been used to establish probiotic potential of the strains. Due to rapid development of next generation sequencing in the past decade, scientists are now exploring whole genome sequencing to identify and functionally characterize novel probiotic strains. A number of potential probiotic strains have been identified based on their putative functional genes from their whole genome [3].

Paenibacillus spp. are well known as growth promoters for crops, but there are very limited studies exploring their probiotic potentials, despite the fact that they can produce diverse kinds of active substances namely antimicrobial peptides (bacteriocin and lipopeptide), volatile organic compounds, and digestive enzymes (amylase, cellulase, lipase, protease, etc), among many others [4].

Previously, we have isolated a novel *Paenibacillus* strain SK3146^T (=KACC 18876^T=LMG 29568^T) from a pig feed, which was taxonomically assigned as *Paenibacillus konkukensis* sp. nov. [5]. In this study, we provide a detailed description of the complete genome sequence of SK3146 and analyzed its putative functional genes related to digestive enzymes and bacteriocin which could be beneficial attributes as a functional feed additive.

SK3146 was cultured in Luria-Bertani broth for two days at 37°C under shaking conditions (100 rpm). Genomic DNA of SK3146 was extracted using the Wizard Genomic DNA Purification Kit (Promega Corporation, Madison, WI, USA) according to the manufacturer's instructions. The genome of SK3146 was completely sequenced using the PacBio® RS II system by Macrogen Inc. (Seoul, South Korea). The PacBio RS II system libraries were prepared using the SMRTbell template prep kit v 1.0. In total, 142,242 bp PacBio subreads with 1,041,901,553 bp were generated using the PacBio® RS II system, and their mean length and N50 value were 7,324 and 10,697 bp, respectively. The sequencing reads were *de novo* assembled using the HGAP analysis with default options. The assembly was completed with the PacBio RS II system. Annotation of coding DNA sequence (CDS) and functional genes were analyzed by the Prokka v1.10. The general features of the SK3146 were

analyzed based on its complete genome sequence using the Geneious 8.1.9 software (Biomatters, New Zealand) [6]. The predicted CDS were classified depending on the clusters of orthologous genes (COG), followed by the construction of a circular genome map and analysis of protein function by a web server: Bacterial Annotation System (<https://www.basys.ca/>) [7].

The circular genome visualization and general features of SK3146 genome are presented in Fig. 1 and summarized in Table 1, respectively. The complete genome of the strain consists of a single circular chromosome measuring 7,968,964 bp in size and 53.4% in G+C content. A total of 6,988 genes were predicted in the genome including 6,842 CDS, 37 ribosomal RNAs (rRNA), 108 transfer RNAs (tRNA), and 1 transfer-messenger RNA (tmRNA) loci. In addition, 10 CRISPR elements, three prophage regions, and 16 IS elements were identified in the genome.

Furthermore, we have analyzed the presence of potential enzymes in the genome of SK3146 via protein function annotation with the Kyoto Encyclopedia of Genes and Genomes database, which provide specific substrates, reactions, and enzyme nomenclature [8]. We then categorized the genes encoding potential enzymes of SK3146 according to enzyme code number. The class of hydrolases (EC3) including galactosidase, glucosidase, cellulase, lipase, xylanase, protease, and others in SK3146 are listed in Table 2. EC3 hydrolases including phosphatases, glycosidases, peptidases, nucleosidases, and lipases are widely used in feed additive industry to improve digestibility and bioavailability of nutrients in animal feeds [9]. Besides enzyme encoding genes, one bacteriocin encoding gene was also found on the chromosome of SK3146. The hydrolytic enzymes such as glucanase, cellulase, protease, and chitinase of *Paenibacillus* have been reported to have anti-fungal activities via destruction of fungal cell wall [10]. Moreover, β -glucosidase, cellulase, xylanase, and protease have been demonstrated to reduce carbohydrate- and protein-based anti-nutritional factors present in the plant-derived protein sources and consequently improving the nutritional quality of feed [11].

In the present study, the complete genome of *P. konkukensis* sp. nov., SK3146 isolated from a pig feed has been reported. The genome of SK3146 encodes multiple enzymes that could be applied to improve the digestibility and bioavailability of nutrients of animal feed. A gene encoding bacteriocin

was also identified. Thus, the genome mining conducted in this study suggests that strain SK3146 has significant potential as a probiotic for use in feed additive applications. In addition, the genome information of SK3146 widens our understanding on the whole genus of *Paenibacillus* to explore and develop next generation probiotics. The genome-based protein prediction will be validated by *in vitro* characterization and *in vivo* animal study in near future.

Genomic sequence accession number

The complete genome sequence of *P. konkukensis* sp. nov., SK3146 was deposited in the GenBank under the accession number **CP027059**.

ACCEPTED

References

1. Barba-Vidal E, Martin-Orue SM, Castillejos L. Practical aspects of the use of probiotics in pig production: A review. *Livest Sci.* 2019;223:84-96.
2. Chaucheyras-Durand F, Durand H. Probiotics in animal nutrition and health. *Benef Microbes.* 2010;1(1):3-9.
3. Cunningham M, Azcarate-Peril MA, Barnard A, Benoit V, Grimaldi R, Guyonnet D, et al. Shaping the Future of Probiotics and Prebiotics. *Trends Microbiol.* 2021.
4. Grady EN, MacDonald J, Liu L, Richman A, Yuan ZC. Current knowledge and perspectives of *Paenibacillus*: a review. *Microb Cell Fact.* 2016;15(1):203.
5. Im WT, Yi KJ, Lee SS, Moon HI, Jeon CO, Kim DW, et al. *Paenibacillus konkukensis* sp. nov., isolated from animal feed. *Int J Syst Evol Microbiol.* 2017;67(7):2343-8.
6. Kears M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics.* 2012;28(12):1647-9.
7. Van Domselaar GH, Stothard P, Shrivastava S, Cruz JA, Guo A, Dong X, et al. BASys: a web server for automated bacterial genome annotation. *Nucleic Acid Res.* 2005;33(Web Server issue):W455-9.
8. McDonald AG, Tipton KF. Fifty-five years of enzyme classification: advances and difficulties. *FEBS J.* 2014;281(2):583-92.
9. Aureli R, La-Marta J, Grossi AB, Della Pia EA, Esteve-Garcia E, Wulf-Andersen L, et al. A novel glucuronoxylan hydrolase produced by fermentation is safe as feed additive: toxicology and tolerance in broiler chickens. *Regul Toxicol Pharmacol.* 2018;99:213-24.
10. Naing KW, Anees M, Kim SJ, Nam Y, Kim YC, Kim KY. Characterization of antifungal activity of *Paenibacillus ehimensis* KWN38 against soilborne phytopathogenic fungi belonging to various taxonomic groups. *Ann Microbiol.* 2014;64(1):55-63.
11. Chi C-H, Cho S-J. Improvement of bioactivity of soybean meal by solid-state fermentation with *Bacillus amyloliquefaciens* versus *Lactobacillus* spp. and *Saccharomyces cerevisiae*. *LWT - Food Sci Technol.*

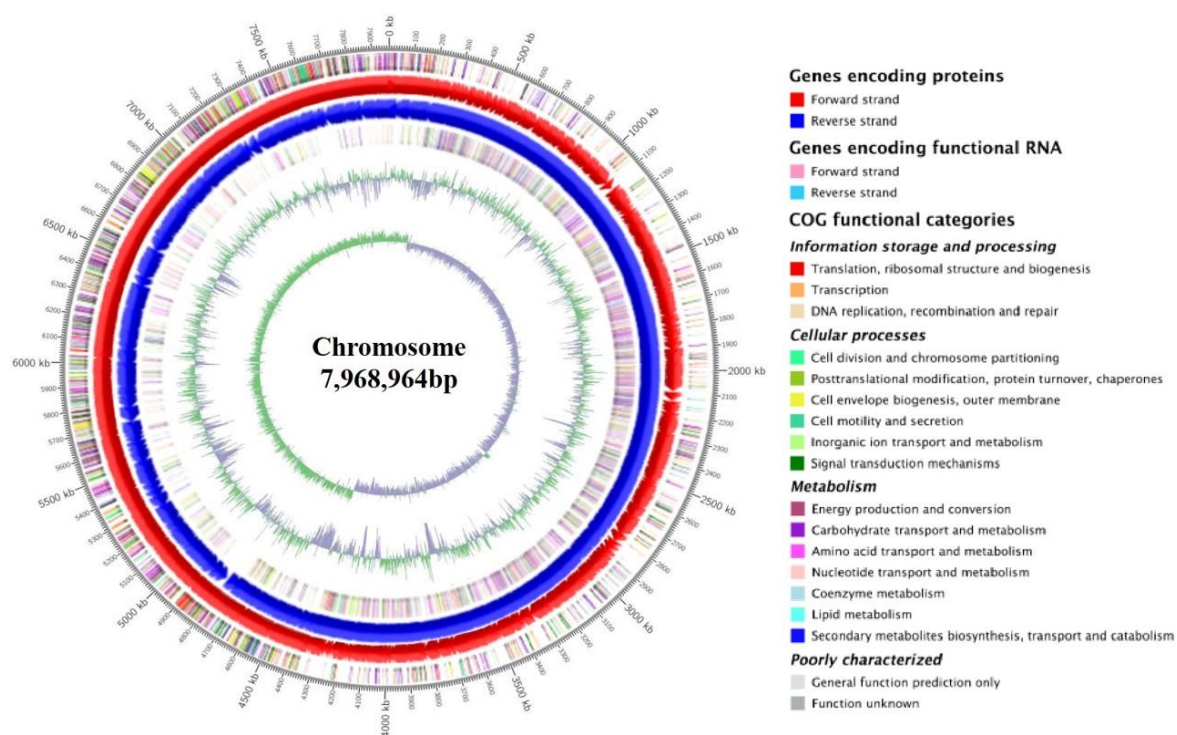


Fig. 1. Circular genome map of *Paenibacillus konkukensis* sp. nov., SK3146. Marked characteristics are shown from outside to the center: COG annotation gene distribution on the forward strand; CDS on the forward DNA strand (red); CDS on the reverse DNA strand (blue); COG annotation gene distribution on the reverse strand; GC content, and GC skew. The lines in each concentric circle indicate the position of the represented feature; the color key is presented on the right side of the map.

Table 1. General features of *Paenibacillus konkukensis* SK3146 genome

Attributes	Value
Length of sequence (bp)	7,968,964
G+C content (%)	53.4
No. of CDS	6,842
No. of genes	6,988
% coding	86
No. of RNAs	146
No. of rRNAs	37
No. of tRNAs	108
No. of tmRNA	1
No. of CRISPR regions	10
No. of prophages	3
No. of IS	16

ACCEPTED

Table 2. Potential metabolic enzyme determinants in *Paenibacillus konkukensis* SK3146 genome

Exoenzymes	No. of CDS	Product Name	EC Number *
Arylsulfatase	38	Arylsulfatase	3.1.
Arabinofuranohydrolase	2	Non-reducing end beta-L-arabinofuranosidase, Arabinoxylan arabinofuranohydrolase precursor	3.2.
Allantoinase	1	Allantoinase	3.5
Agmatinase	1	Agmatinase	3.5.
Arginase	1	Arginase	3.5.
α -galactosidase	5	Alpha-galactosidase	3.2.
α -galacturonidase	1	Alpha-galacturonidase	3.2.
β -glucuronidase	6	Beta-glucuronidase	3.2.
β -glucosidase	4	Periplasmic beta-glucosidase precursor, Thermostable beta-glucosidase B, Oligo-1%2C6-glucosidase	3.2.
β -galactosidase	8	Beta-galactosidase, Evolved beta-galactosidase subunit alpha	3.2.
β -xylosidase	3	Beta-xylosidase	3.2.
Cellulase	7	Endoglucanase precursor, Endoglucanase Z precursor	3.2.
Chitinase	1	Chitinase A1 precursor	3.2.
D-aminoacylase	1	D-aminoacylase	3.5.
Galactose	3	Galactose/methyl galactoside import ATP-binding protein	3.6.
Glutaminase	1	Glutaminase	3.5.
Glucoamylase	3	Glucoamylase precursor	3.2.
Lipase	1	Lipase 3 precursor	3.1.
Nitrilase	1	Nitrilase	3.5.
Protease	25	Germination protease precursor, Protease 3 precursor, Putative metalloprotease, Putative zinc protease, Serine endoprotease, Serine protease Do-like, Serine protease, Rhomboid protease, Carboxy-terminal processing protease precursor, putative CtpA-like serine protease, ATP-dependent protease subunit, putative protease precursor, Sporulation-specific protease, Protease, ATP-dependent protease proteolytic subunit, Lon protease, putative protease, Putative cysteine protease	3.4 3.2.
Polygalacturonase	3	Polygalacturonase	3.2.
Thiaminase	2	Thiaminase-2	3.5.
Urease	6	Urease subunit beta, Urease subunit gamma, Urease subunit alpha	3.5.
Xylanase	27	Endo-1%2C4-beta-xylanase A precursor, Endo-1%2C4-beta-xylanase Z precursor, Beta-1%2C3-xylanase XYL4 precursor	3.2.

*EC (Enzyme commission number) is a numerical classification scheme for enzymes, based on the chemical reactions they catalyze. EC 3 enzymes are hydrolases (EC3.1: ester bonds; EC3.2: sugars; EC3.3: ether bonds; EC3.4: Peptide bonds; EC3.5: carbon-nitrogen bonds; EC3.6: acid anhydrides).