J. Microbiol. Biotechnol. (2015), 25(12), 1971–1976
http://dx.doi.org/10.4014/jmb.1507.07077

jmb

Genomic Analysis of the Moderately Haloalkaliphilic Bacterium *Oceanobacillus kimchii* Strain X50^T with Improved High-Quality Draft Genome Sequences

Dong-Wook Hyun, Tae Woong Whon, Joon-Yong Kim, Pil Soo Kim, Na-Ri Shin, Min-Soo Kim, and Jin-Woo Bae*

Department of Life and Nanopharmaceutical Sciences and Department of Biology, Kyung Hee University, Seoul 02453, Republic of Korea

Received: July 22, 2015
Revised: September 11, 2015
Accepted: September 14, 2015

First published online
September 15, 2015

**Corresponding author*
Phone: +82-2-961-2312;
Fax: +82-2-961-0244;
E-mail: baejw@khu.ac.kr

pISSN 1017-7825, eISSN 1738-8872

Oceanobacillus kimchii is a member of the genus *Oceanobacillus* within the family Bacillaceae. Species of the *Oceanobacillus* possess moderate haloalkaliphilic features and originate from various alkali or salty environments. The haloalkaliphilic characteristics of *Oceanobacillus* advocate they may have possible uses in biotechnological and industrial applications, such as alkaline enzyme production and biodegradation. This study presents the draft genome sequence of *O. kimchii* X50^T and its annotation. Furthermore, comparative genomic analysis of *O. kimchii* X50^T was performed with two previously reported *Oceanobacillus* genome sequences. The 3,822,411 base-pair genome contains 3,792 protein-coding genes and 80 RNA genes with an average G+C content of 35.18 mol%. The strain carried 67 and 13 predicted genes annotated with transport system and osmoregulation, respectively, which support the tolerance phenotype of the strain in high-alkali and high-salt environments.

Keywords: Moderately halophile, alkaliphile, *Oceanobacillus*, *Oceanobacillus kimchii*, Bacillaceae

Introduction

The genus *Oceanobacillus*, a member of family Bacillaceae, was first introduced by Lu *et al*. (2001) [18]. At the time of writing, the genus comprises 17 validated species and two subspecies. Members of the genus *Oceanobacillus* are gram-stain-positive, motile, and endospore-forming rods. Most species in the genus *Oceanobacillus* are characterized as moderate haloalkaliphilic organisms and have been found in salty or alkaline environments, such as fermented indigo [7, 8], fermented food [20, 30], and marine environments [6, 15, 18].

Moderately alkaliphilic bacteria can survive in alkali environment in the pH 9–10 range [11]. To adapt to high external pH, these bacteria have homeostasis mechanisms for neutralizing cytoplasmic pH, such as Na^+/H^+ antiporter-dependent pH homeostasis [12]. Moderately halophilic bacteria can grow in salty condition within the range of 5–20% (w/v) NaCl by regulating their osmotic concentrations [1]. As a consequence of these osmoregulation strategies, these bacteria can use osmolytes or compatible solutes, such as betaines, polyols, and ectoines, under high-salt environmental conditions [5, 22]. Collectively, the haloalkaliphilic features of *Oceanobacillus* species imply that these organisms may have biotechnological applications such as for organic pollutants biodegradation and alternative energy production [14] and alkaline enzymes [25].

Oceanobacillus kimchii type strain X50^T (= DSM 23341^T = JCM 16803^T = KCTC 14914^T) was isolated from a traditional Korean fermented food known as "mustard kimchi" [30]. The strain X50^T showed the same haloalkaliphilic features as those of other *Oceanobacillus* species and can grow in 0–15% (w/v) NaCl and pH 7.0–10, and shows optimal growth at pH 9 [30]. The present study summarizes the polyphasic features of *O. kimchii* X50^T, and provides not only genomic information derived from its draft genome sequence but also compares the features of strain X50^T with those of other *Oceanobacillus* species.

Materials and Methods

Phylogenetic Analysis Based on 16S rRNA Gene Sequences

The taxonomic position of *O. kimchii* X50^T was confirmed, based on its sequence of 16S rRNA gene. Comparison of the 16S rRNA gene sequence between strain *O. kimchii* X50^T and closely related type strains in the EzTaxon-e database [9] indicated that strain X50^T is a member of the genus *Oceanobacillus* in the family Bacillaceae. The strain shares 98.86% sequence similarity with *O. iheyensis* HTE831^T and 96.09% sequence similarity with invalid species *O. massiliensis* N'Diop^T. A phylogenetic consensus tree was constructed to determine the phylogenetic relationships between strain X50^T and other *Oceanobacillus* species. The 16S rRNA gene sequences of strain X50^T and other *Oceanobacillus* species were aligned using the multiple sequence alignment program Clustal W [29]. Phylogenetic tree construction was performed using aligned sequences with maximum-likelihood [3], maximum-parsimony [10], and neighbor-joining [24] algorithms applying 1,000 bootstrap replicates by MEGA 6 [27].

Genomic DNA Extraction, Sequencing, and Sequence Assembly

The biomass of *O. kimchii* strain X50^T was prepared at 30ºC for 2 days in 1% (w/v) NaCl-containing marine 2216 medium (Difco). Genomic DNA extraction was performed with a Wizard Genomic DNA Purification Kit (Promega A1120). Three platforms were used for DNA sequencing: an Illumina Hiseq system with a 150 base pair (bp) paired end library, a 454 Genome Sequencer FLX Titanium system (Roche Diagnostics) with an 8 kb paired end library, and a PacBio *RS* system (Pacific Biosciences) by ChunLab Inc., Korea. The sequencing reads assemblies were carried out using CLCbio CLC Genomics Workbench 5.0 (CLCbio) and Roche gsAssembler 2.6 (Roche Diagnostics). In the process of sequences assembly, sequencing reads acquired from the 454 Genome Sequencer FLX Titanium system and Illumina Hiseq system were primary integrated, and then sequencing reads acquired from PacBio *RS* system were used for gap filling. The genome project is deposited in the Genomes OnLine Database [16] and the genome sequence is deposited in GenBank. A summary of the project information is shown in Table 1.

Gene Prediction and Annotation

The open reading frames (ORFs) were predicted by the Integrated Microbial Genomes-Expert Review (IMG-ER) pipeline [19]. Gene annotation and functional comparisons of the predicted ORFs were conducted using the IMG-ER platform [19] with NCBI COG [28], NCBI Refseq [21], and Pfam [4] databases. GLIMMER 3.02 [2] was used for the gene calling method. IMG-ER platform [19], tRNAscan-SE 1.23 [17], and RNAmer 1.2 [13] were utilized to find tRNA genes and rRNA genes.

Genomic Sequence and Functional Profile Comparison

Average nucleotide identities (ANI) between the three *Oceanobacillus* genome sequences were calculated by using the Ez-Taxon-e server. Functional profile-based correlation values were calculated by using the IMG-ER platform with COG, Pfam, KO, and TIGRfam profiles.

Results

The phylogenetic analysis indicated that *O. kimchii* X50^T fell into a clade in the genus *Oceanobacillus* and formed a cluster with *O. iheyensis*, which is the closest related species to *O. kimchii* (Fig. 1).

A total of 6,523,431 sequencing reads (267.1-fold genome coverage) were obtained using a combination of the Ilumina Hiseq system (6,376,362 reads; 251.9-fold coverage), Roche 454 system (130,352 reads; 5.6-fold coverage), and PacBio

Table 1. Summary of genome sequencing information.

Property	Term
Sequencing platforms	Illumina Hiseq, 454 GS FLX Titanium, and PacBio *RS* system
Sequencing libraries	Illumina library and 454 PE library (8 kb insert size)
Number of reads	6,523,431 sequencing reads
Finishing quality	Improved high-quality draft
Coverage	267.1-fold coverage (5.6 × 454 pyrosequencing, 251.9 × Illumina, and 9.6 × PacBio)
Assemblers	gsAssembler 2.6, CLC Genomics Workbench 5.0
Scaffolds	1
Contigs	20
Gene calling method	GLIMMER 3.02
GenBank accession number	AOCX01000000
NCBI ID	175944
IMG-ER number	2528311005
Source material identifier	KCTC 14914^T, DSM 23341^T, JCM 16803^T

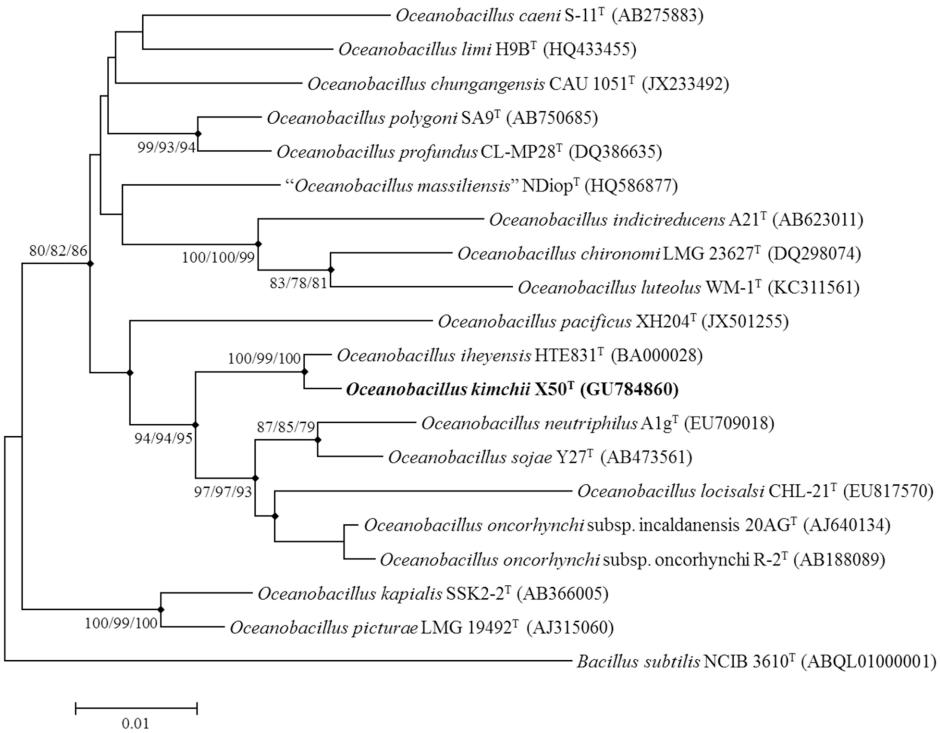

Fig. 1. Phylogenetic consensus tree based on 16S rRNA gene sequences showing the relationship between *Oceanobacillus kimchii* X50^T and the related type strains of *Oceanobacillus* species.

Bacillus subtilis subsp. *subtilis* NCIB 3610^T was set as an outgroup. Filled diamonds represent identical branches revealed in the phylogenetic consensus trees constructed using the neighbor-joining, maximum-likelihood, and maximum-parsimony algorithms. The GenBank accession numbers for the 16S rRNA genes of each strain are shown in parentheses. The numbers at the nodes indicate the bootstrap values as percentages of 1,000 replicates. The scale bar represents 0.01 accumulated changes per nucleotide.

RS system (16,717 reads; 9.6-fold coverage). The assembled genome sequence of *O. kimchii* strain X50^T comprises a single scaffold that includes 20 contigs and contains 3,822,411 bp with 35.18 mol% G+C content.

The genome has the capacity to code for a total of 3,872 predicted genes. Of these, 3,792 were assigned to protein-coding genes (97.93%) and 80 were assigned to RNA genes (2.07%), including 48 tRNA genes and 10 rRNA genes (two 16S rRNA, four 5S rRNA, and four 23S rRNA genes). A total of 3,193 predicted genes (82.46%) were assigned to have putative functions, whereas the remaining genes (17.54%) were considered as hypothetical proteins. Moreover, 2,635 genes were classified under 23 COG functional categories. The overall genome statistics are summarized in Table 2 and visualized in Fig. 2. The gene distributions in the COG functional categories are shown in Table 3.

The genome of *O. kimchii* X50^T contains 19 predicted genes associated with antibiotic resistance, including genes coding for resistance to vancomycin, fosfomycin, and beta-

Table 2. Genome statistics.

Attribute	Value	% of total
Genome size (bp)	3,822,411	100.00%[a]
DNA coding region (bp)	3,287,166	86.00%[a]
DNA G+C content (bp)	1,344,750	35.18%[a]
Predicted genes	3,872	100.00%[b]
RNAs	80	2.07%[b]
Protein-coding genes	3,792	97.93%[b]
Genes with predicted functions	3,193	82.46%[b]
Genes with enzyme	1,014	26.19%[b]
Genes assigned to COGs	2,635	68.05%[b]
Genes assigned Pfam domains	3,278	84.66%[b]
Genes with signal peptides	218	5.63%[b]
Genes with transmembrane helices	1,105	28.54%[b]
Genes in biosynthetic clusters	229	5.91%[b]
Fused protein-coding gene	77	1.99%[b]

The percent of total is based on either [a]the genome size (bp) or [b]the total gene number in the annotated genome.

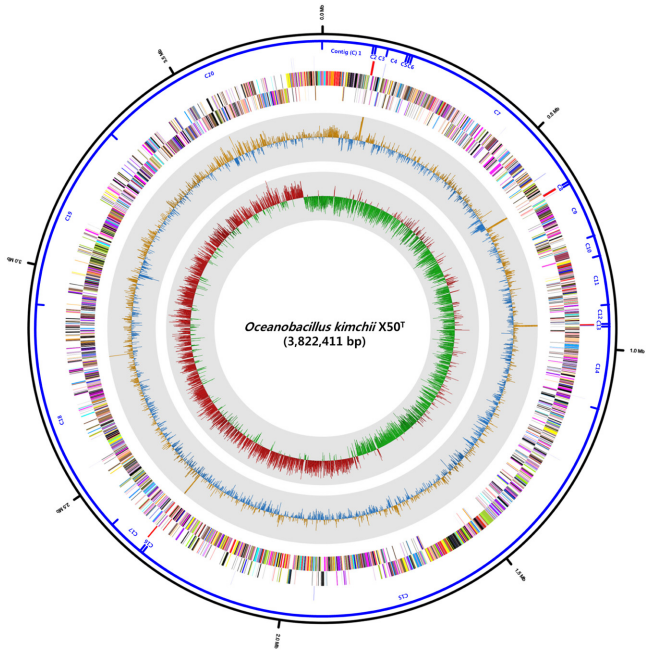

Fig. 2. Graphical circular genome map.
From the outside to the inside: contigs on pseudochromosome (blue circle), RNA genes (red, tRNAs; blue, rRNAs), genes on the forward strand (colored following to COG categories), and genes on the reverse strand (colored following to COG categories). The inner circle shows the GC skew, where yellow indicates positive values and blue indicates negative values. The GC ratio is shown in red and green, which indicate positive and negative, respectively. Gaps between individual contigs are not presented.

lactams, such as vancomycin B-type resistance protein (VanW) and beta-lactamase class A. Strain X50^T encodes 96 predicted genes associated with sporulation, such as spore germination protein (GerKA) and stage V sporulation protein (SpoVAB). Sixty-nine motility-associated genes were predicted, including 12 chemotaxis and 57 flagella genes. These genes are characteristic of bacteria that engage in sporulation and flagella motility, which is the case for *O. kimchii*. To identify prophages in *O. kimchii* X50^T, PHAST [31] was used. One prophage was identified in a

Table 3. Numbers of genes annotated with the 23 general COG functional categories.

Code	Value	% of total[a]	Description
G	275	9.15	Carbohydrate metabolism and transport
E	271	9.02	Amino acid metabolism and transport
K	237	7.88	Transcription
J	211	7.02	Translation
P	176	5.85	Inorganic ion metabolism and transport
H	159	5.29	Coenzyme metabolism and transport
C	156	5.19	Energy production and conversion
M	149	4.96	Cell-wall/membrane biogenesis
T	133	4.42	Signal transduction mechanisms
I	132	4.39	Lipid metabolism and transport
L	108	3.59	Replication, recombination, and repair
O	100	3.33	Posttranslational modification, protein turnover, and chaperones
F	99	3.29	Nucleotide metabolism and transport
Q	79	2.63	Secondary metabolites biosynthesis, catabolism, and transport
V	70	2.33	Defense mechanisms
N	52	1.73	Cell motility
D	48	1.60	Cell cycle control, mitosis, and meiosis
U	28	0.93	Intracellular trafficking and secretion
X	12	0.40	Mobilome: prophages, transposons
W	2	0.07	Extracellular structures
B	1	0.03	Chromatin structure and dynamics
R	308	10.25	General function prediction only
S	200	6.65	Unknown function
-	1,237	31.95	Not assigned in COGs

[a]The number of protein-coding genes in the annotated genome was considered as the total, which was used for proportion calculation.

Table 4. Average nucleotide identity values and correlation scores.

Species	1	2	3
1. *O. kimchii*	-	88.86%[a] (0.98/0.96/0.94/0.94)[b]	69.51%[a] (0.93/0.88/0.87/0.82)[b]
2. *O. iheyensis*	88.72%[a] (0.98/0.96/0.94/0.94)[b]	-	69.34%[a] (0.93/0.88/0.86/0.83)[b]
3. *O. massiliensis*	69.45%[a] (0.93/0.88/0.87/0.82)[b]	69.66%[a] (0.93/0.88/0.86/0.83)[b]	-

[a]Average nucleotide identity value (%).
[b]Pearson coefficient (Pfam/KO/TIGRfam/COG).

20.8 kb (with 36.5% G+C content) region containing 26 CDS. Fifteen of the 26 CDS were annotated with *Bacillus* phage protein. *O. kimchii* X50^T encodes 13 predicted genes associated with osmotic stress regulation, such as choline-glycine betaine transporter (BetT) and periplasmic glycine betaine/choline-binding lipoprotein of the ABC-type transport system (OpuAA). Strain X50^T possesses 67 predicted genes associated with membrane transporter systems, including ABC transporters, Na^+/H^+ antiporter, protein translocation systems, cation transporters, and TRAP transporters, such as the C4-dicarboxylate transport system. These genes could be key factors allowing *O. kimchii* X50^T to adapt to high-salt and high-alkali environments *via* osmotic regulation and adjustment of cytoplasmic pH, respectively.

To date, two *Oceanobacillus* strains, *O. iheyensis* HTE831^T and *O. massiliensis* N'diop^T, have been sequenced and validated [23, 26]. *O. kimchii* X50^T has the largest genome and highest number of predicted genes, but has the lowest G+C content of the validated genomes of the two *Oceanobacillus* strains. *O. kimchii* X50^T showed 88.86% (88.72% in reciprocal) and 69.51% (69.45% in reciprocal) ANI with *O. iheyensis* HTE831^T and *O. massiliensis* N'diop^T, respectively (Table 4). *O. kimchii* X50^T has 0.94 to 0.98 correlation values (Pearson coefficient) with *O. iheyensis* HTE831^T and 0.82 to 0.93 correlation values with *O. massiliensis* N'diop^T. The results of genome sequence and functional profile analyses indicate that *O. kimchii* X50^T shares more genomic and functional features with *O. iheyensis* HTE831^T than with *O. massiliensis* N'diop^T.

Discussion

At the time of writing, *O. kimchii* X50^T is the third strain in the genus *Oceanobacillus* to be subjected to genome analysis; the other two are *O. iheyensis* and *O. massiliensis*. The comparative analyses, which were based on average nucleotide genomic sequence similarities and correlations of functional profiles, show that *O. kimchii* X50^T is more closely related to *O. iheyensis* HTE831^T than to *O. massiliensis* N'diop^T, which is in line with the results of 16S rRNA gene sequence-based phylogenetic analysis. The *O. kimchii* X50^T genome encodes sporulation, flagella motility, osmoregulation, and pH homeostasis genes in accordance with the previously reported characteristics of *O. kimchii*. Further studies are required to elucidate the mechanisms involved in the osmoregulation and pH homeostasis of this haloalkaliphilic bacterium. The results of such analyses could facilitate its use in biotechnological applications.

Acknowledgments

This work was carried out with the support of "the Next-Generation Biogreen21 Program (Project No. PJ008208012014)", the Rural Development Administration, and "the Strategic Initiative for Microbiomes in Agriculture and Food (Project No. 914006-4-SB010)", Ministry of Agriculture, Food and Rural Affairs, Republic of Korea.

References

1. DasSarma S, DasSarma P. 2001. Halophiles. *In: Encyclopedia of Life Sciences* (*ELS*). John Wiley & Sons, Ltd., NY.
2. Delcher AL, Bratke KA, Powers EC, Salzberg SL. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* **23:** 673-679.
3. Felsenstein J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* **17:** 368-376.
4. Finn RD, Mistry J, Tate J, Coggill P, Heger A, Pollington JE, *et al.* 2010. The Pfam protein families database. *Nucleic Acids Res.* **38:** D211-D222.
5. Galinski EA. 1993. Compatible solutes of halophilic eubacteria – molecular principles, water-solute interaction, stress protection. *Experientia* **49:** 487-496.
6. Graham RL, Pollock CE, O'Loughlin SN, Ternan NG, Weatherly DB, Tarleton RL, McMullan G. 2007. Multidimensional analysis of the insoluble sub-proteome of *Oceanobacillus iheyensis* HTE831, an alkaliphilic and halotolerant deep-sea bacterium isolated from the Iheya ridge. *Proteomics* **7:** 82-91.
7. Hirota K, Aino K, Nodasaka Y, Yumoto I. 2013. *Oceanobacillus indicireducens* sp. nov., a facultative alkaliphile that reduces an indigo dye. *Int. J. Syst. Evol. Microbiol.* **63:** 1437-1442.
8. Hirota K, Hanaoka Y, Nodasaka Y, Yumoto I. 2013. *Oceanobacillus polygoni* sp. nov., a facultatively alkaliphile isolated from indigo fermentation fluid. *Int. J. Syst. Evol. Microbiol.* **63:** 3307-3312.
9. Kim OS, Cho YJ, Lee K, Yoon SH, Kim M, Na H, *et al.* 2012. Introducing EzTaxon-e: a prokaryotic 16S rRNA gene sequence database with phylotypes that represent uncultured species. *Int. J. Syst. Evol. Microbiol.* **62:** 716-721.
10. Kluge AG, Farris JS. 1969. Quantitative phyletics and the evolution of anurans. *Syst. Zool.* **18:** 1-32.
11. Krulwich TA, Ito M. 2013. Alkaliphilic Prokaryotes, pp. 441-469. *In* Rosenberg E, DeLong E, Lory S, Stackebrandt E, Thompson F (eds.). *The Prokaryotes*. Springer, Berlin–Heidelberg.
12. Krulwich TA, Sachs G, Padan E. 2011. Molecular aspects of bacterial pH sensing and homeostasis. *Nat. Rev. Microbiol.* **9:** 330-343.
13. Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW. 2007. RNAmmer: consistent and rapid

annotation of ribosomal RNA genes. *Nucleic Acids Res.* **35:** 3100-3108.
14. Le Borgne S, Paniagua D, Vazquez-Duhalt R. 2008. Biodegradation of organic pollutants by halophilic bacteria and archaea. *J. Mol. Microbiol. Biotechnol.* **15:** 74-92.
15. Lee SY, Oh TK, Kim W, Yoon JH. 2010. *Oceanobacillus locisalsi* sp. nov., isolated from a marine solar saltern. *Int. J. Syst. Evol. Microbiol.* **60:** 2758-2762.
16. Liolios K, Chen IM, Mavromatis K, Tavernarakis N, Hugenholtz P, Markowitz VM, Kyrpides NC. 2010. The Genomes On Line Database (GOLD) in 2009: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res.* **38:** D346-D354.
17. Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25:** 955-964.
18. Lu J, Nogi Y, Takami H. 2001. *Oceanobacillus iheyensis* gen. nov., sp. nov., a deep-sea extremely halotolerant and alkaliphilic species isolated from a depth of 1050 m on the Iheya Ridge. *FEMS Microbiol. Lett.* **205:** 291-297.
19. Markowitz VM, Mavromatis K, Ivanova NN, Chen IM, Chu K, Kyrpides NC. 2009. IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics* **25:** 2271-2278.
20. Namwong S, Tanasupawat S, Lee KC, Lee JS. 2009. *Oceanobacillus kapialis* sp. nov., from fermented shrimp paste in Thailand. *Int. J. Syst. Evol. Microbiol.* **59:** 2254-2259.
21. Pruitt KD, Tatusova T, Brown GR, Maglott DR. 2012. NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Res.* **40:** D130-D135.
22. Roberts MF. 2005. Organic compatible solutes of halotolerant and halophilic microorganisms. *Saline Syst.* **1:** 5.
23. Roux V, Million M, Robert C, Magne A, Raoult D. 2013. Non-contiguous finished genome sequence and description of *Oceanobacillus massiliensis* sp. nov. *Stand. Genomic Sci.* **9:** 370-384.
24. Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4:** 406-425.
25. Sarethy IP, Saxena Y, Kapoor A, Sharma M, Sharma SK, Gupta V, Gupta S. 2011. Alkaliphilic bacteria: applications in industrial biotechnology. *J. Ind. Microbiol. Biotechnol.* **38:** 769-790.
26. Takami H, Takaki Y, Uchiyama I. 2002. Genome sequence of *Oceanobacillus iheyensis* isolated from the Iheya Ridge and its unexpected adaptive capabilities to extreme environments. *Nucleic Acids Res.* **30:** 3927-3935.
27. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* **30:** 2725-2729.
28. Tatusov RL, Galperin MY, Natale DA, Koonin EV. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* **28:** 33-36.
29. Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22:** 4673-4680.
30. Whon TW, Jung MJ, Roh SW, Nam YD, Park EJ, Shin KS, Bae JW. 2010. *Oceanobacillus kimchii* sp. nov. isolated from a traditional Korean fermented food. *J. Microbiol.* **48:** 862-866.
31. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. 2011. PHAST: a fast phage search tool. *Nucleic Acids Res.* **39:** W347-W352.