



ELSEVIER

Contents lists available at ScienceDirect

## Data in Brief

journal homepage: [www.elsevier.com/locate/dib](http://www.elsevier.com/locate/dib)

## Data Article

# Draft genome sequencing data of the moderately halophilic bacterium, *Allobacillus halotolerans* SKP2-8 from shrimp paste (*ka-pi*)



Engkarat Kingkaew<sup>a</sup>, Supalurk Yiamsombut<sup>a</sup>, Saranporn Poothong<sup>a</sup>,  
Wenyu Shi<sup>b</sup>, Linhuan Wu<sup>b</sup>, Juncai Ma<sup>b</sup>, Somboon Tanasupawat<sup>a,\*</sup>

<sup>a</sup> Department of Biochemistry and Microbiology, Faculty of Pharmaceutical Sciences, Chulalongkorn University, Bangkok 10330, Thailand

<sup>b</sup> World Data Center for Microorganisms (WDCM), Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, PR China

## ARTICLE INFO

## Article history:

Received 23 June 2022

Revised 18 July 2022

Accepted 17 August 2022

Available online 22 August 2022

Dataset link: [Allobacillus sp. SKP2-8, whole genome shotgun sequencing project \(Original data\)](#)

## Keywords:

*Allobacillus*

Draft genome

Fermented food

Moderately halophile

Shrimp paste

## ABSTRACT

A moderately halophilic, Gram-stain-positive, spore-forming rod-shaped bacterium, designated SKP2-8 was isolated from a traditional fermented shrimp paste (*Ka-pi*) collected from the market in Samut Sakhon province, Thailand. This isolate SKP2-8 was closely related to *Allobacillus halotolerans* LMG 24826<sup>T</sup> with 99.56% similarity based on 16S rRNA gene sequence. The draft genome of SKP2-8 was 2.53 Mb with 2,515 coding sequences with an average G+C content of 39.5 mol%. The ANiB, ANiM, AAI and the digital DNA-DNA hybridization values of isolate SKP2-8 were 97.22%, 97.64%, 97.75% and 78.0%, respectively, compared with *A. halotolerans* LMG 24826<sup>T</sup>. Based on the phenotypic characteristics, DNA-DNA relatedness and phylogenomic analysis, it was identified as *Allobacillus halotolerans*. The genome sequence data of this isolate provide information for further analysis of the potential biotechnological use of this microorganism and guide the characterization. The draft genome was deposited at DDBJ/EMBL/GenBank (DNA Databank of Japan/European Molecular Biology Laboratory/Genbank) (VMHF00000000).

\* Corresponding author.

E-mail address: [somboon.t@chula.ac.th](mailto:somboon.t@chula.ac.th) (S. Tanasupawat).

## Specifications Table

Subject	Biology
Specific subject area	Microbiology, Genomics, Biotechnology
Type of data	Table Figure Excel Sheets Datasets (Word)
How the data were acquired	SEM, Illumina Miseq, RAST annotation, DFAST annotation, PATRIC annotation, PathogenFinder, PlasmidFinder, The Comprehensive Antibiotic Resistance Database
Data format	Analyzed and deposited
Description of data collection	<i>Allobacillus</i> sp. SKP2-8 was cultivated on JCM No.377 medium. Genomic DNA was extracted from a pure culture of <i>Allobacillus</i> sp. SKP2-8. The obtained sequencing data were used for genome analysis, identification, and search of stress tolerance and lipase-, esterase- associated genes.
Data source location	City/Town/Region: Samut Sakhon Country: Thailand Shrimp paste (ka-pi)
Data accessibility	<ul style="list-style-type: none"> <li>- The whole genomes were deposited at DDBJ/EMBL/GenBank under the accession number Isolate SKP2-8 (VMHF000000000) (<a href="https://www.ncbi.nlm.nih.gov/nucleotide/VMHF000000000.1/">https://www.ncbi.nlm.nih.gov/nucleotide/VMHF000000000.1/</a>), and isolate LMG 24856<sup>T</sup> (JAHLF000000000) (<a href="https://www.ncbi.nlm.nih.gov/nucleotide/JAHLF000000000.1/">https://www.ncbi.nlm.nih.gov/nucleotide/JAHLF000000000.1/</a>)</li> <li>- The description characteristics of isolate SKP2-8, (<a href="https://doi.org/10.6084/m9.figshare.20124134.v1">https://doi.org/10.6084/m9.figshare.20124134.v1</a>)</li> <li>- The lipase activity (<i>in vitro</i>) analysis and lipase gene, (<a href="https://doi.org/10.6084/m9.figshare.20128685.v2">https://doi.org/10.6084/m9.figshare.20128685.v2</a>)</li> <li>- The scanning electron micrograph (SEM) of isolate SKP2-8 and LMG 24826<sup>T</sup> (<a href="https://doi.org/10.6084/m9.figshare.20116040.v1">https://doi.org/10.6084/m9.figshare.20116040.v1</a>)</li> <li>- The safety and pathogenicity assessment (<a href="https://doi.org/10.6084/m9.figshare.20117243.v2">https://doi.org/10.6084/m9.figshare.20117243.v2</a>)</li> <li>- The stress tolerance and lipase-, esterase- associated genes (<a href="https://doi.org/10.6084/m9.figshare.20116103.v1">https://doi.org/10.6084/m9.figshare.20116103.v1</a>)</li> </ul>
Related research article	Yiamsombut S, Kanchanasin P, Phongsoptanun W, Kuncharoen N, Savarajara A, Shi W, Wu L, Ma J, Tanasupawat S. <i>Allobacillus salarius</i> sp. nov., and <i>Allobacillus saliphilus</i> sp. nov., isolated from shrimp paste (ka-pi) in Thailand. Arch Microbiol. 2021 Dec 24;204(1):71. doi:10.1007/s00203-021-02694-9. PMID: 34951663.

## Value of the Data

- These data provide the source for the description of *Allobacillus halotolerans* that was originally published only single isolate.
- These data are fundamental to environmental and clinical microbiology.
- These data serve to conduct comparative genomics in moderate-halotolerant related gene and allow a better understanding of the mechanisms involved in osmotic stress.
- This study provides the genome analysis of lipase and esterase genes.

## 1. Data Description

The genus *Allobacillus*, a moderately halophilic rod-shaped, isolated from shrimp paste in Taiwan, Republic of China, was proposed by Sheu et al., and *Allobacillus halotolerans* was the type

**Table 1**Genomic features of *Allobacillus halotolerans* SKP2-8 and *Allobacillus halotolerans* LMG 24826<sup>T</sup>.

Attribute	SK2-8	LMG 24826 <sup>T</sup>
Isolation source	Shrimp paste ( <i>ka-pi</i> )	Shrimp paste
Genbank accession	VMHF00000000	JAHLZF000000000
Biosample accession	SAMN12329002	SAMN19700765
Bioproject accession	PRJNA555754	PRJNA737595
Genome size (bp)	2,533,751	2,726,708
G+C content (%)	39.5	39.5
Genome coverage	500x	300x
Total genes	2,580	2,854
Total CDss	2,515	2,776
Total proteins	2,475	2,744
rRNA	2, 1, 1 (5S, 16S, 23S)	4, 6, 7 (5S, 16S, 23S)
tRNA	57	57
ncRNA	4	4
N50	58,702	104,182
L50	14	8
Contig	49	96

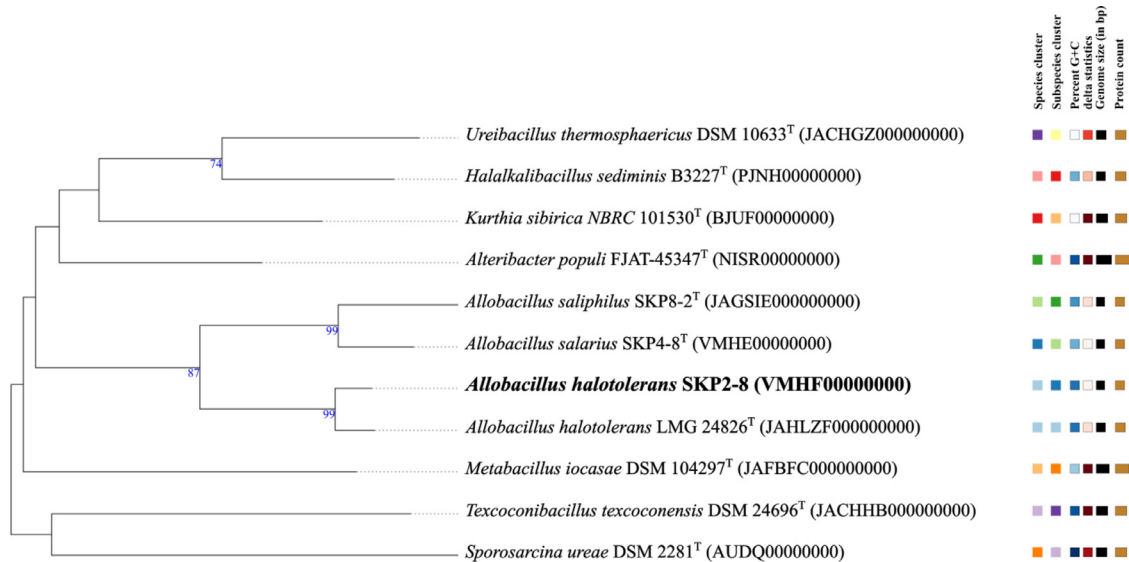
species [1]. *Allobacillus salarius* and *Allobacillus saliphilus* isolated from shrimp paste (Ka-pi) in Thailand are proposed as the second and third species [2].

The description of characteristics of isolate SKP2-8 is described in supplementary file 1 (<https://doi.org/10.6084/m9.figshare.20124134.v1>) and the scanning electron micrograph (SEM) of isolate SKP2-8 and LMG 24826<sup>T</sup> is shown in supplementary file 2 (<https://doi.org/10.6084/m9.figshare.20116040.v1>) [2]. In addition, the Table 1 described the results of the genomic features of SKP2-8 and LMG 24826<sup>T</sup>. The draft genome sequence of isolate SKP2-8 was 2,533,751 bp, with a genomic G+C content of 2,515 coding sequences, 65 RNAs and genome coverage of 500×. The total number of genes after annotation was 2,580, of which 2,475 were coding sequences, 3 were ribosomal RNAs, and 57 tRNAs.

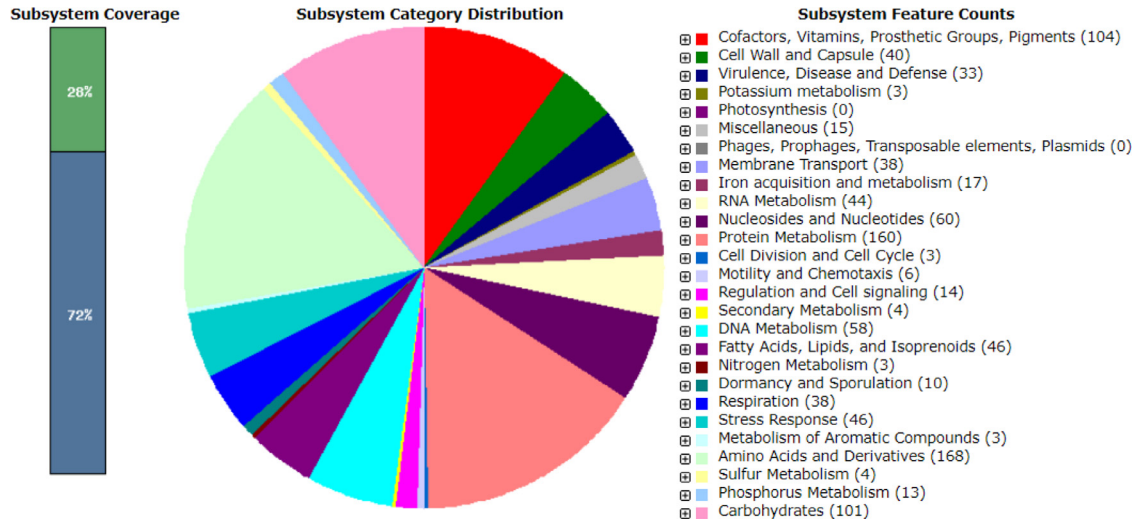
Based on full 16S rRNA gene sequence, the isolate SKP2-8 (1,464 bp) was closely related to *A. halotolerans* LMG 24826<sup>T</sup> with 99.56% similarity. The phylogenomic analysis demonstrated the cluster formation of isolate SKP2-8 with the *A. halotolerans* LMG 24826<sup>T</sup> (Fig. 1).

The ANiB and ANiM values of the draft genomes between isolate SKP2-8 and *A. halotolerans* LMG 24826<sup>T</sup> were 97.22 and 97.64%, respectively. The average amino acid identity (AAI) value between isolate SKP2-8 and *A. halotolerans* LMG 24826<sup>T</sup> was 97.75% (from 2,446 proteins), [3]. In addition, the dDDH value of the draft genome between isolate SKP2-8 and *A. halotolerans* LMG 24826<sup>T</sup>, was 78% (C.I. 75 – 80.7%) [4].

The safety and pathogenicity evaluation of isolate SKP2-8 and LMG 24828<sup>T</sup> are shown in supplementary file 3 (<https://doi.org/10.6084/m9.figshare.20117243.v2>). From the RAST (Fig. 2), DFAST, and PATRIC annotation, the isolate SKP 2-8 and LMG 24826<sup>T</sup> contained halotolerant-associated and lipase/esterase-associated genes and they was described in supplementary file 4 (<https://doi.org/10.6084/m9.figshare.20116103.v1>) which are responsible for osmotic-stress response, fatty acids, lipids, and isoprenoids metabolism [5–8]. Furthermore, the analysis of the lipase activity (*in vitro*) and *in silico* analysis is shown in supplementary file 5 (<https://doi.org/10.6084/m9.figshare.20128685.v2>) [8].



**Fig. 1.** Phylogenomic tree based on whole genome sequence data result of SKP2-8 and closely related type strain reconstructed on the Type (Strain) Genome Server (TYGS).



**Fig. 2.** Subsystem distribution of *Allobacillus halotolerans* SKP2-8 constructed from the RAST annotation server.

## 2. Experimental Design, Materials and Methods

*Allobacillus halotolerans* SKP2-8 was isolated from shrimp paste collected from the market in Samut Sakhon province, Thailand by using spread-plate technique duplicate on modified JCM medium no.377 agar plates. Lipase activity was screened and determined [8]. Whole genome sequence was performed using an Illumina Miseq platform (Illumina, Inc., San Diego, US-CA) by the World Data Center for Microorganisms (WDCM) under the Global Catalogue of Microorganisms (GCM) 2.0 project. Assembling the reads to contigs were accomplished by using SPAdes 3.12 [9]. The genomic quality was qualified by CheckM [10]. The genome was annotated by using the DFAST sever [11], Rapid Annotation Server Technology (RAST) [12], PATRIC [13], the NCBI Prokaryotic Genome Annotation Pipeline (PGAP). A phylogenetic tree based on whole-genome sequence was constructed by using TYGS web server (<https://tygs.dsmz.de/>) [14]. Antibiotic resistance genes were determined using the Comprehensive Antibiotic Resistance Database (CARD; <https://card.mcmaster.ca>) [15]. The pathogenicity was predicted by PathogenFinder web-based tool [16] and plasmid was detected by PlasmidFinder [17]. Average nucleotide identity (ANI) values were calculated with pairwise genome alignment of the draft genome sequences of *Allobacillus halotolerans* LMG 24826<sup>T</sup> (JAHLZF000000000) by using the ANI-BLAST (ANiB) and ANI-MUMmer (ANIm) algorithms [18] implemented within the JspeciesWS web service [19]. The average amino acid identity (AAI) was calculated by web-based (<http://enve-omics.ce.gatech.edu/aai/>) [3]. Calculation of the digital DNA-DNA hybridization (dDDH) values was achieved by using the Genome-to-Genome Distance Calculator (GGDC 2.1) using the BLAST+ method [20]. Results were based on the recommended formula 2 (identities/HSP length), which is useful when dealing with incomplete draft genomes.

### Ethics Statements

No ethical issue.

### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Data Availability

*Allobacillus* sp. SKP2-8, whole genome shotgun sequencing project (Original data) (NCBI).

### CRediT Author Statement

**Engkarat Kingkaew:** Conceptualization, Methodology, Validation, Formal analysis, Data curation, Writing – original draft, Visualization; **Supalurk Yiamsombut:** Methodology, Validation, Formal analysis, Investigation, Resources; **Saranporn Poothong:** Conceptualization, Methodology, Validation, Formal analysis, Data curation, Writing – original draft, Visualization; **Wenyu Shi:** Software, Formal analysis, Resources, Project administration; **Linhuan Wu:** Software, Formal analysis, Resources; **Juncai Ma:** Software, Formal analysis, Resources; **Somboon Tanasupawat:** Conceptualization, Validation, Resources, Data curation, Writing – review & editing, Supervision, Project administration, Funding acquisition.

## Acknowledgments

This study was supported by the Grant for International Research Integration: Research Pyramid, Ratchadaphiseksomphot Endowment Fund (CUGRP-61-01-33-01), Chulalongkorn University and the International Partnership Program of Chinese Academy of Sciences (Grant No. 153211KYSB 201900211).

## References

- [1] S.-Y. Sheu, A.B. Arun, S.-R. Jiang, C.-C. Young, W.-M. Chen, *Allobacillus halotolerans* gen. nov., sp. nov. isolated from shrimp paste, *Int. J. Syst. Evol. Microbiol.* 61 (5) (2011) 1023–1027.
- [2] S. Yiamsombut, P. Kanchanasin, W. Phongsopitanun, N. Kuncharoen, A. Savarajara, W. Shi, L. Wu, J. Ma, S. Tanasupawat, *Allobacillus salarius* sp. nov., and *Allobacillus saliphilus* sp. nov., isolated from shrimp paste (ka-pi) in Thailand, *Arch. Microbiol.* 204 (1) (2022) 1–6.
- [3] L.M. Rodriguez-R, K.T. Konstantinidis, Bypassing cultivation to identify bacterial species, *Microbe* 9 (3) (2014) 111–118.
- [4] S.W. Ahn, S.H. Lee, H.-S. Son, S.W. Roh, Y.-E. Choi, Genomic analysis of halophilic bacterium, *Lentibacillus* sp. CBA3610, derived from human feces, *Gut Pathog.* 13 (1) (2021) 41, doi:10.1186/s13099-021-00436-2.
- [5] G.J. Gregory, E.F. Boyd, Stressed out: Bacterial response to high salinity using compatible solute biosynthesis and uptake systems, lessons from *Vibrionaceae*, *Comput. Struct. Biotechnol.* 19 (2021) 1014–1027, doi:10.1016/j.csbj.2021.01.030.
- [6] L. Matarredona, M. Camacho, B. Zafrilla, M.-J. Bonete, J. Esclapez, The role of stress proteins in haloarchaea and their adaptive response to environmental shifts, *Biomolecules* 10 (10) (2020) 1390.
- [7] Y.-H. Chen, Y.-T. Shyu, S.-S. Lin, Characterization of candidate genes involved in halotolerance using high-throughput omics in the halotolerant bacterium *Virgibacillus chiguensis*, *PLoS One* 13 (8) (2018) e0201346–e0201346, doi:10.1371/journal.pone.0201346.
- [8] F. Cardenas, E. Alvarez, M.-S. de Castro-Alvarez, J.-M. Sanchez-Montero, M. Valmaseda, S.W. Elson, J.-V. Sinisterra, Screening and catalytic activity in organic synthesis of novel fungal and yeast lipases, *J. Mol. Catal. B Enzym.* 14 (4–6) (2001) 111–123.
- [9] A. Bankevich, S. Nurk, D. Antipov, A.A. Gurevich, M. Dvorkin, A.S. Kulikov, V.M. Lesin, S.I. Nikolenko, S. Pham, A.D. Prjibelski, SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing, *J. Comput. Biol.* 19 (5) (2012) 455–477.
- [10] D.H. Parks, M. Imelfort, C.T. Skennerton, P. Hugenholtz, G.W. Tyson, CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes, *Genome Res.* 25 (7) (2015) 1043–1055.
- [11] Y. Tanizawa, T. Fujisawa, Y. Nakamura, DFAST: a flexible prokaryotic genome annotation pipeline for faster genome publication, *Bioinformatics* 34 (6) (2018) 1037–1039.
- [12] R.K. Aziz, D. Bartels, A.A. Best, M. DeJongh, T. Disz, R.A. Edwards, K. Formsma, S. Gerdes, E.M. Glass, M. Kubal, The RAST Server: rapid annotations using subsystems technology, *BMC Genom.* 9 (1) (2008) 1–15.
- [13] J.J. Davis, A.R. Wattam, R.K. Aziz, T. Brettin, R. Butler, R.M. Butler, P. Chlenski, N. Conrad, A. Dickerman, E.M. Dietrich, The PATRIC bioinformatics resource center: expanding data and analysis capabilities, *Nucleic Acids Res.* 48 (D1) (2020) D606–D612.
- [14] J.P. Meier-Kolthoff, M. Göker, TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy, *Nat. Commun.* 10 (1) (2019) 1–10.
- [15] B.P. Alcock, A.R. Raphenya, T.T.Y. Lau, K.K. Tsang, M. Bouchard, A. Edalatmand, W. Huynh, A.-L.V. Nguyen, A.A. Cheng, S. Liu, CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database, *Nucleic Acids Res* 48 (D1) (2020) D517–D525.
- [16] S. Cosentino, M. Voldby Larsen, F. Møller Aarestrup, O. Lund, PathogenFinder-distinguishing friend from foe using bacterial whole genome sequence data, *PLoS one* 8 (10) (2013) e77302.
- [17] A. Carattoli, E. Zankari, A. Garcia-Fernandez, M.V. Larsen, O. Lund, L. Villa, F.M. Aarestrup, H. Hasman, PlasmidFinder and pMLST: *in silico* detection and typing of plasmids, *Antimicrob. Agents Chemother.* 58 (7) (2014) 3895–3903.
- [18] M. Richter, R. Rosselló-Móra, Shifting the genomic gold standard for the prokaryotic species definition, *Proc. Natl. Acad. Sci.* 106 (45) (2009) 19126–19131.
- [19] M. Richter, R. Rosselló-Móra, F. Oliver Glöckner, J. Peplis, JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison, *Bioinformatics* 32 (6) (2016) 929–931.
- [20] J.P. Meier-Kolthoff, A.F. Auch, H.-P. Klenk, M. Göker, Genome sequence-based species delimitation with confidence intervals and improved distance functions, *BMC Bioinformatics* 14 (1) (2013) 1–14.