





Bacillus paralicheniformis isolated from a hypersaline sediment of Lake Texcoco, draft genome sequence of strain TXO7B-1SG6

Lorna Catalina Can-Ubando,^a  Ayixon Sánchez-Reyes,^b Gaudy Lizeth Manzaneres-Leal,^a Keila Isaac-Olivé,^c Horacio Sandoval-Trujillo,^d  Ninfa Ramírez-Durán^a

^aLaboratorio de Investigación en Microbiología Médica y Ambiental, Facultad de Medicina, Universidad Autónoma del Estado de México, Toluca, México, México

^bInvestigador por México, CONACYT—Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, México

^cLaboratorio de Investigación en Teranóstica, Facultad de Medicina, Universidad Autónoma del Estado de México, Toluca, México, México

^dLaboratorio de Producción de Biológicos, Departamento de Sistemas Biológicos, Universidad Autónoma Metropolitana—Xochimilco, Ciudad de México, México

ABSTRACT We present the draft genome sequence of the halotolerant strain *Bacillus paralicheniformis* TXO7B-1SG6.

B *Bacillus paralicheniformis*, a Gram-positive bacillus cultured in medium with up to 10% NaCl, was described as a new species in 2015 (1). It controls plant pathogens and produces bioactive secondary metabolites (2). We present the draft genome sequence of strain TXO7B-1SG6, isolated from hypersaline sediment from Lake Texcoco, Mexico (19°28'42"N, 98°58'26"W), in June 2019 and identified as *Bacillus paralicheniformis*.

Sediment solutions (0.1 mg/mL concentration) were prepared for isolation. The dilutions were inoculated onto petri dishes using Czapek culture medium and brain heart infusion (BHI) agar (37°C, 7 to 10 days). The strain was isolated and cultured on halophilic medium (HM) (3), being halotolerant (NaCl growth range, 0 to 5%; optimum, 5%) with a pH range of 7 to 10 (optimum 8).

DNA was extracted using the Promega Wizard genomic kit (A1125). The 16S rRNA gene was amplified by PCR with the universal primers 27F and 1492R (4) and sequenced by Psomagen (MD, USA). A consensus sequence (1,436 bp) was obtained using the BioEdit v7.09 program. BLAST (5) and the EzBioCloud database (6) were used for sequence comparison. The highest identity percentages were found with *Bacillus haynesii* (97.63%; BLAST) and *Bacillus paralicheniformis* (98.52%; EzBioCloud).

To sequence the TXO7B-1SG6 genome, the strain was inoculated onto HM agar to obtain biomass. High-molecular-weight (HMW) DNA was extracted using the Quick-DNA HMW MagBead kit (Zymo Research catalog number D6060). The sample was subjected to long-read sequencing using the MinION platform (ONT) of the Institute of Biotechnology at the National Autonomous University of Mexico (IBT-UNAM).

A MinION library was prepared from unshered genomic DNA using the ligation sequencing kit (SQK-LSK109). The EXP-NBD104 barcoding kit was used for multiplexing and the R9.4.1 flow cell for sequencing. The reads were base called and demultiplexed using Guppy v4.4.1 (7). Raw read quality exploration using NanoPlot showed 138,694 reads with a mean length of 1,608.8 nucleotides (nt) and an N_{50} value of 12,283 nt. The reads were quality filtered using Filtlong v0.2.1 (8), and adapters were trimmed using Porechop v0.2.4 (9). A total of 138,694 reads were obtained with a mean read quality of 12.5. The filtered reads were assembled using Unicycler v0.4.8 software (option *-l long-read-only assembly*) (10), and the final genome was rotated (the *dnaA* sequence was set to the first nucleotide position) using the Circlator v1.5.5 pipeline (11). The genome was annotated using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) v6.1 (12). Default parameters were used for all software unless otherwise noted.

A chromosome-level assembly was achieved, resulting in 1 single replicon with a size of

Editor Steven R. Gill, University of Rochester School of Medicine and Dentistry

Copyright © 2022 Can-Ubando et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Ninfa Ramírez-Durán, ninfard@hotmail.com.

The authors declare no conflict of interest.

Received 4 July 2022

Accepted 13 July 2022

4,523,282 bp, a 45.60% G+C content, and an average read sequencing coverage depth of 50×. The completeness and contamination were estimated using CheckM v1.1.1 (13) at 94.5% and 0.3%, respectively. The average nucleotide identity (ANI) was estimated according to the NCBI workflow (14). The genome of TXO7B-15G6 has an ANI of 98.72% to *B. paralicheniformis* strain KJ-16 (GenBank accession number [GCA_001042485.2](https://www.ncbi.nlm.nih.gov/nuccore/GCA_001042485.2)).

A total of 4,159 coding genes were identified, including 4 noncoding RNAs (ncRNAs), 24 rRNAs (8 5S, 8 16S, and 8 23S), and 82 tRNAs.

Data availability. The consensus sequence of the 16S rRNA gene has been deposited at GenBank under accession number [OM746865](https://www.ncbi.nlm.nih.gov/nuccore/OM746865). The draft genome sequence has been deposited at NCBI GenBank under accession number [CP082897](https://www.ncbi.nlm.nih.gov/nuccore/CP082897); the version described in this article is version [CP082897.3](https://www.ncbi.nlm.nih.gov/nuccore/CP082897.3). The associated BioSample accession number is [SAMN20057199](https://www.ncbi.nlm.nih.gov/biosample/SAMN20057199), and the BioProject accession number is [PRJNA743662](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA743662). The raw reads were deposited at the SRA under accession number [SRR18345719](https://www.ncbi.nlm.nih.gov/sra/SRR18345719).

ACKNOWLEDGMENTS

This work is derived from the project “Análisis genómico de bacterias halotolerantes degradadoras de contaminantes emergentes de origen farmacéutico: diclofenaco e ibuprofeno,” corresponding to a specialized research stay granted within the framework of the call Estancias de Investigación Especializadas COMECYT-EDOMEX Segunda promoción 2021.

We are grateful for the financial support of the Secretaría de Investigación y Estudios Avanzados, Universidad Autónoma del Estado de México through the research project “Análisis de genomas bacterianos de *Bacillus* sp. con potencial degradador de los antiinflamatorios no esteroideos: Diclofenaco e Ibuprofeno” (6543/2022CIB).

This work results from a collaboration established through the Academic and Collaborative Network Microbiología y Química en Ciencias de la Salud (1039/2014RIFC).

This research did not receive any specific funding from any public, commercial, or nonprofit funding agency.

REFERENCES

- Dunlap CA, Kwon SW, Rooney AP, Kim SJ. 2015. *Bacillus paralicheniformis* sp. nov., isolated from fermented soybean paste. *Int J Syst Evol Microbiol* 65:3487–3492. <https://doi.org/10.1099/ijsem.0.000441>.
- Daas MS, Rosana ARR, Acedo JZ, Douzane M, Nateche F, Kebbouche-Gana S, Vederas JC. 2018. Draft genome sequence of *Bacillus paralicheniformis* F47, isolated from an Algerian salty lake. *Genome Announc* 6:e00190-18. <https://doi.org/10.1128/genomeA.00190-18>.
- Quesada E, Ventosa A, Rodríguez-Valera F, Megias L, Ramos-Cormenzana A. 1983. Numerical taxonomy of moderately halophilic Gram-negative bacteria from hypersaline soils. *J Gen Microbiol* 129:2649–2657.
- Frank JA, Reich CI, Sharma S, Weisbaum JS, Wilson BA, Olsen GJ. 2008. Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes. *Appl Environ Microbiol* 74:2461–2470. <https://doi.org/10.1128/AEM.02272-07>.
- Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. 2013. GenBank. *Nucleic Acids Res* 41:D36–D42. <https://doi.org/10.1093/nar/gks1195>.
- Yoon SH, Ha S, Kwon S, Lim J, Kim Y, Seo H, Chun J. 2017. Introducing EzBioCloud: a taxonomically united database of 16S rRNA and whole genome assemblies. *Int J Syst Evol Microbiol* 67:1613–1317. <https://doi.org/10.1099/ijsem.0.001755>.
- Oxford Nanopore Technologies. 2022. Guppy v4.4.1. <https://nanoporetech.com/nanopore-sequencing-data-analysis>.
- Wick RR. 2021. Filtlong v0.2.1. Github. <https://github.com/rwrick/Filtlong>.
- Wick RR, Volkening J. 2018. Porechop v0.2.4. <https://github.com/rwrick/Porechop>.
- Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 13:e1005595. <https://doi.org/10.1371/journal.pcbi.1005595>.
- Hunt M, Silva ND, Otto TD, Parkhill J, Keane JA, Harris SR. 2015. Circlator: automated circularization of genome assemblies using long sequencing reads. *Genome Biol* 16:294. <https://doi.org/10.1186/s13059-015-0849-0>.
- Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. *Nucleic Acids Res* 44:6614–6624. <https://doi.org/10.1093/nar/gkw569>.
- Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 25:1043–1055. <https://doi.org/10.1101/gr.186072.114>.
- Ciufu S, Kannan S, Sharma S, Badretdin A, Clark K, Turner S, Brover S, Schoch CL, Kimchi A, DiCuccio M. 2018. Using average nucleotide identity to improve taxonomic assignments in prokaryotic genomes at the NCBI. *Int J Syst Evol Microbiol* 68:2386–2392. <https://doi.org/10.1099/ijsem.0.002809>.