AMERICAN SOCIETY FOR MICROBIOLOGY

Microbiology® Resource Announcements

GENOME SEQUENCES

Draft Genome Sequence of *Bacillus pacificus* KVCMST-8A-12, Isolated from a Marine Sediment Sample from the Kanyakumari Coast, India

Asha Santhi,[a] Venkatesh Subramanian,[b] Krishnaveni Muthan[a]

[a]Immuno-Pharmacology Laboratory, Center for Marine Science and Technology, Manonmaniam Sundaranar University, Marina Campus, Rajakkamangalam, Kanyakumari District, Tamil Nadu, India

[b]Genetic Engineering and Regenerative Biology Laboratory, Department of Biotechnology, Manonmaniam Sundaranar University, Tirunelveli, Tamil Nadu, India

ABSTRACT A DNase-producing *Bacillus pacificus* strain was isolated, and the whole-genome sequence is reported in this paper. The draft genome sequence of *Bacillus pacificus* KVCMST-8A-12 constitutes 2.4 Gbp of raw reads, with a GC content of 35.24%. In total, 5,661 protein-coding genes, 64 tRNA genes, and 4 rRNA genes were predicted.

A spore-forming, rod-shaped, aerobic bacterium designated *Bacillus pacificus* KVCMST-8A-12 was isolated from a sediment sample from the coastal region in Kanyakumari, India (1). This strain was reported to produce extracellular DNase, which can be commercially exploited (1). The whole genome of *B. pacificus* KVCMST-8A-12 was sequenced. Briefly, genomic DNA was extracted and purified from *B. pacificus* KVCMST-8A-12, which had been grown at 32°C in Zobell marine broth, using a QIAmp DNA extraction kit (Qiagen) and following the manufacturer's instructions. DNA quantity and quality were assessed using NanoDrop measurements at 260 nm and 280 nm and resolution in a 0.8% agarose gel, respectively. The library was prepared following the workflow of the TruSeq DNA library preparation kit (Illumina) (2). The library was sequenced with a high-throughput sequencer (Illumina HiSeq 2500 sequencer) using paired-end sequencing strategies with 150-bp paired-end sequencing.

A total of 2.4 Gb of raw data was generated from 8,053,402 raw reads for *B. pacificus* KVCMST-8A-12; the reads were assembled using SPAdes v3.14.1 (3) and FastQC v0.11.8 (4) after filtering of low-quality reads (Q scores of $\leq$5) and adaptor sequences using Trimmomatic v0.38 (5). Reads were polished using Pilon v1.23 and Unicycler v0.4.9 (6). A genome size of 5,508,258 bp annotated with 383 scaffolds and 411 contigs (N_{50} value of 219,973 bp), with a GC content of 35.24% and 100× coverage, was obtained for our strain. The online RAST gene annotation pipeline (7) predicted 5,661 protein-coding genes. The annotation included 1,474 hypothetical proteins and 4,187 proteins with functional assignments. The proteins with functional assignments included 1,093 proteins with Enzyme Commission (EC) numbers (8), 920 proteins with Gene Ontology (GO) assignments (9), and 810 proteins mapped to KEGG pathways (10). PATRIC v3.6.12 annotation includes two types of protein families (11, 12). This genome has 5,547 proteins that belong to cross-genus protein families (PGFams). The average protein length determined by Prokka was 276 amino acids, and 28 rRNAs, 1 transfer-messenger RNA, and 77 tRNAs were obtained through Prokaryotic Genome Annotation Pipeline (PGAP) annotation. When annotation with Prokka v1.12 (13) and annotation with NCBI PGAP (14) were compared with RAST annotation, a few differences were noted. This may be due to different parameters and databases used by these three tools. Further annotation is in progress to identify the genes responsible for DNase production, which could find utility in various bioscience fields. Default parameters were used for all software unless otherwise noted.

Editor Frank J. Stewart, Montana State University

Address correspondence to Krishnaveni Muthan, dr.krishnavenimuthan@gmail.com.

The authors declare no conflict of interest.

Received 13 October 2021
Accepted 8 November 2021
Published 16 December 2021

Data availability. The whole-genome sequence of *B. pacificus* KVCMST-8A-12 has been deposited in GenBank under BioProject accession number PRJNA656513 with BioSample accession number SAMN16200492, and raw reads can be found under SRA accession number SRX9149524.

ACKNOWLEDGMENTS

A.S. was supported by an INSPIRE fellowship from the Department of Science and Technology, India (grant IF140069). Samples used in this study were collected under a project funded by the University Grants Commission, New Delhi, India [project 45-572/2012 (SR), dated 18 July 2012].

The author contributions were as follows: A.S., conceptualization, data curation, formal analysis, investigation, and methodology; V.S., data curation, validation, and writing (review and editing); K.M., conceptualization, supervision, and writing (original draft).

REFERENCES

1. Asha S, Krishnaveni M. 2020. Isolation and molecular level identification of DNase producing halophilic *Bacillus cereus* family isolates from marine sediment sample. J Pure Appl Microbiol 14:423–435. https://doi.org/10.22207/JPAM.14.1.44.
2. Yan L, Yang M, Guo H, Yang L, Wu J, Li R, Liu P, Lian Y, Zheng X, Yan J, Huang J, Li M, Wu X, Wen L, Lao K, Li R, Qiao J, Tang F. 2013. Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. Nat Struct Mol Biol 20:1131–1139. https://doi.org/10.1038/nsmb.2660.
3. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.
4. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc.
5. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
6. 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.
7. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST Server: Rapid Annotations using Subsystems Technology. BMC Genomics 9:75. https://doi.org/10.1186/1471-2164-9-75.
8. Schomburg I, Chang A, Ebeling C, Gremse M, Heldt C, Huhn G, Schomburg D. 2004. BRENDA, the enzyme database: updates and major new developments. Nucleic Acids Res 32:D431–D433. https://doi.org/10.1093/nar/gkh081.
9. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. 2000. Gene ontology: tool for the unification of biology. Nat Genet 25:25–29. https://doi.org/10.1038/75556.
10. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2016. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res 44:D457–D462. https://doi.org/10.1093/nar/gkv1070.
11. Wattam AR, Davis JJ, Assaf R, Boisvert S, Brettin T, Bun C, Conrad N, Dietrich EM, Disz T, Gabbard JL, Gerdes S, Henry CS, Kenyon RW, Machi D, Mao C, Nordberg EK, Olsen GJ, Murphy-Olson DE, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Vonstein V, Warren A, Xia F, Yoo H, Stevens RL. 2017. Improvements to PATRIC, the all-bacterial bioinformatics database and analysis resource center. Nucleic Acids Res 45:D535–D542. https://doi.org/10.1093/nar/gkw1017.
12. Wattam AR, Brettin T, Davis JJ, Gerdes S, Kenyon R, Machi D, Mao C, Olson R, Overbeek R, Pusch GD, Shukla MP, Stevens R, Vonstein V, Warren A, Xia F, Yoo H. 2018. Assembly, annotation, and comparative genomics in PATRIC, the all bacterial bioinformatics resource center. Methods Mol Biol 1704:79–101. https://doi.org/10.1007/978-1-4939-7463-4_4.
13. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153.
14. 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.