

Complete genome sequence of the calcite-precipitating bacterium *Cytobacillus kochii* strain ITBMC_116, isolated from a limestone karst in Vietnam

Thiet M. Vu,¹ Loan Q. Le,² Thanh M. Luc,¹ Khoa H. D. Do,¹ Danh H. Nguyen,¹ Kien T. Tran,^{2,3} Nhung T. T. Vu,² Dung H. Nguyen^{2,3}

AUTHOR AFFILIATIONS See affiliation list on p. 2.

ABSTRACT We report the complete genome of calcite-precipitating *Cytobacillus kochii* strain ITBMC_116, isolated from limestone karst in Vietnam. Hybrid Illumina–Nanopore sequencing yielded a 4.83-Mbp chromosome and an 87.4-kbp plasmid. This resource supports studies of microbially induced calcite precipitation and biomineralization-based biotechnologies.

KEYWORDS *Cytobacillus kochii*, microbially induced calcite precipitation, whole genome sequencing

Bacteria within the genus *Cytobacillus* are of increasing biotechnological interest due to their potential for microbially induced calcite precipitation (1). This biomineralization capability is critical for sustainable civil engineering applications, such as biocementation, soil stabilization, and the development of self-healing concrete (2). The isolation of novel, robust calcite-precipitating bacteria from unique environments provides valuable genetic resources for understanding and harnessing these capabilities. Here, we present the complete genome sequence of *Cytobacillus kochii* strain ITBMC_116, isolated from a limestone sample collected in the karst region of Ho Chi Minh City, Vietnam (geographical coordinates: 10.544574, 107.166390). The strain was deposited at the Department of Applied Microbiology, Institute of Life Sciences (contact: Dung H. Nguyen, email: nhdung@ils.vast.vn).

A single colony of strain ITBMC_116 was inoculated into 2 mL of Nutrient Broth medium (Oxoid) and cultured overnight (14 h) at 30°C with shaking. Total genomic DNA was extracted from the cell pellet using a DNA Miniprep Kit (BioBasic) and size-selected using magnetic beads (Beckman Coulter). The quality and quantity of the DNA sample were checked using agarose gel (1%) electrophoresis and Qubit 4 fluorometer (Thermo Fisher Scientific), respectively. Short-read sequencing was performed by KTest Science Co. Ltd. (Ho Chi Minh City, Vietnam) on an Illumina MiSeq platform (2 × 150 bp paired-end), generating 44,772,776 reads (~6.7 Gbp of data). Long-read sequencing was conducted in-house using an Oxford Nanopore Technologies (ONT) MinION Mk1B device with Ligation Sequencing Kit (SQK-LSK109) and an R9.4 flow cell, yielding 163,079 reads (~1.87 Gbp of data). For quality control, raw Illumina reads were assessed with FastQC (v0.11.9) and trimmed using fastp (3). ONT reads were base-called with Guppy (v6.5.7) in high-accuracy mode and filtered using NanoPlot (4) to retain reads ≥ 1,000 bp.

A hybrid assembly of the high-quality Illumina (44,480,667 reads) and ONT (103,313 reads) data was performed using Unicycler (v0.5.0) (5) and resolved two complete circular replicons, including a single chromosome (4,830,791 bp) with an average coverage depth of 1,345× and a GC content of 36.8% and plasmid pITBMC_116 (87,365 bp) with an average coverage depth of 1,941× and a GC content of 33%. The assembly

Editor André O. Hudson, Rochester Institute of Technology, Rochester, New York, USA

Address correspondence to Dung H. Nguyen, nhdung@ils.vast.vn.

The authors declare no conflict of interest.

Received 25 November 2025

Accepted 6 January 2026

Published 11 February 2026

Copyright © 2026 Vu 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/).

quality was assessed using CheckM2 (v1.1.0) (6) for contamination (0.21%) and BUSCO (v5.8.0) (7) for completeness (96.9%).

Taxonomic identification using the Type Strain Genome Server (8) and GTDB-Tk (v2.3.2) (9) classified the isolate as *Cytobacillus kochii*, with *C. kochii* DSM 23667^T as the closest reference. Annotation via the NCBI Prokaryotic Genome Annotation Pipeline (PGAP, release 6.1) (10) predicted 5,015 chromosomal genes, including 4,802 protein-coding sequences (CDSs), 107 tRNAs, and 30 rRNAs. The plasmid carried 98 predicted CDS. Default parameters were used for all software.

ACKNOWLEDGMENTS

This research was funded by the Vietnamese Ministry of Science and Technology under grant number ĐTDL.CN-116/21.

We acknowledge Nguyen Tat Thanh University for supporting the high-performance computing system and the ONT sequencing platform.

T.M.V., Conceptualization, Writing – original draft; T.M.L., Formal analysis, Writing – review and editing; L.Q.L., Bacterial isolation and characterization; K.H.D.D., D.H.N., genome sequencing and formal analysis; K.T.T., N.T.T.T., validation, Writing – review and editing; D.H.N., Conceptualization, Funding acquisition, and Writing – review and editing.

AUTHOR AFFILIATIONS

¹Functional Genomics Research Center, NTT Hi-tech Institute, Nguyen Tat Thanh University, Ho Chi Minh City, Vietnam

²Institute of Life Sciences, Vietnam Academy of Science and Technology, Ho Chi Minh City, Vietnam

³Biotechnology Department, Graduate University of Science and Technology, Vietnam Academy of Science and Technology, Hanoi City, Vietnam

AUTHOR ORCIDs

Thiet M. Vu  <http://orcid.org/0000-0002-7950-0758>

Dung H. Nguyen  <http://orcid.org/0000-0002-1108-3521>

AUTHOR CONTRIBUTIONS

Thiet M. Vu, Conceptualization, Funding acquisition, Investigation, Validation, Writing – original draft, Writing – review and editing | Loan Q. Le, Methodology, Project administration, Writing – review and editing | Thanh M. Luc, Data curation, Formal analysis, Investigation, Writing – review and editing | Khoa H. D. Do, Formal analysis, Validation, Writing – review and editing | Danh H. Nguyen, Formal analysis, Methodology, Validation | Kien T. Tran, Validation, Writing – review and editing | Nhung T. T. Vu, Methodology, Validation | Dung H. Nguyen, Conceptualization, Data curation, Funding acquisition, Supervision, Writing – review and editing

DATA AVAILABILITY

This Whole Genome Shotgun project has been deposited at GenBank under the BioProject accession number [PRJNA1310224](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1310224) and BioSample [SAMN50770592](https://www.ncbi.nlm.nih.gov/biosample/SAMN50770592). The complete chromosome and plasmid sequences are available under accession numbers [JBQSLM010000001](https://www.ncbi.nlm.nih.gov/nuccore/JBQSLM010000001) and [JBQSLM010000002](https://www.ncbi.nlm.nih.gov/nuccore/JBQSLM010000002), respectively. The chromosome sequence described in this paper is GenBank accession version [CM126253.1](https://www.ncbi.nlm.nih.gov/nuccore/CM126253.1). The raw Illumina and ONT reads have been deposited in the Sequence Read Archive (SRA) under accession numbers [SRR36067571](https://www.ncbi.nlm.nih.gov/sra/SRR36067571) and [SRR36067570](https://www.ncbi.nlm.nih.gov/sra/SRR36067570), respectively.

REFERENCES

1. Moqsud MA, Gochi T. 2023. Biocementation of slope soil by using native *Cytobacillus horneckiae*. *Bioresource Technology Reports* 23:101520. <https://doi.org/10.1016/j.biteb.2023.101520>
2. Joshi S, Goyal S, Mukherjee A, Reddy MS. 2017. Microbial healing of cracks in concrete: a review. *J Ind Microbiol Biotechnol* 44:1511–1525. <https://doi.org/10.1007/s10295-017-1978-0>
3. Chen S, Zhou Y, Chen Y, Gu J. 2018. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34:i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>
4. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. 2018. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* 34:2666–2669. <https://doi.org/10.1093/bioinformatics/bty149>
5. 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>
6. Chklovski A, Parks DH, Woodcroft BJ, Tyson GW. 2023. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. *Nat Methods* 20:1203–1212. <https://doi.org/10.1038/s41592-023-01940-w>
7. Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva EV, Zdobnov EM. 2018. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol Biol Evol* 35:543–548. <https://doi.org/10.1093/molbev/msx319>
8. Meier-Kolthoff JP, Göker M. 2019. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nat Commun* 10:2182. <https://doi.org/10.1038/s41467-019-10210-3>
9. Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil P-A, Hugenholtz P. 2022. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res* 50:D785–D794. <https://doi.org/10.1093/nar/gkab776>
10. 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>