

Draft genome sequence of epiphytic *Bacillus safensis* strain Cap07D, isolated from *Chenopodium album* L. seeds

Vladimir K. Chebotar,¹ Veronika D. Antonova,² Maksim R. Losev,¹ Maria S. Gancheva,^{1,3} Dmitriy V. Kudryavtcev,¹ Olga A. Bortsova,¹ Elena P. Chizhevskaya,¹ Veronika N. Pishchik¹

AUTHOR AFFILIATIONS See affiliation list on p. 2.

ABSTRACT Here, we report the genome sequence of the epiphytic *Bacillus safensis* Cap07D, a bacterium isolated from *Chenopodium album* L. seeds. The raw sequencing data were obtained using the NextSeq 1000 technology (Illumina).

KEYWORDS epiphytic, *Bacillus safensis*, genome sequence, *Chenopodium album*, seeds

Bacillus safensis L. is a Gram-positive, spore-forming bacterium that belongs to the *Bacillus pumilus* group (1). Members of this group are widely distributed in both marine and terrestrial environments, particularly in soil ecosystems, and are frequently associated with plants as part of their natural microbiota (1). Strain Cap07D was isolated from the seeds of *Chenopodium album* L. collected in 2024 in the Stavropol region of Russia (45.156812° N, 42.098796° E). A flask containing 100 mL of sterile phosphate buffer and 10 g of *C. album* seeds was placed on an ultrasonic bath (Bandelin; 50 Hz) for 10 min. Then, the phosphate buffer solution containing microorganisms washed off from the grains was serially diluted, and 0.1 mL of various dilutions was inoculated on Petri dishes containing LB (Luria Bertani, Sigma-Aldrich, USA) agar medium, and plates were incubated at 28°C for 3 days. A single emerging colony of strain Cap07D was transferred to Luria–Bertani (LB) medium and cultured overnight at 30°C.

Genomic DNA from the strain Cap07D was extracted using the Wizard Genomic DNA Purification Kit (Promega, USA). Library preparation was performed using the KAPA HyperPlus Kit (Roche, Switzerland). Sequencing was conducted on an Illumina NextSeq 1000 platform with the NextSeq 1000/2000 P1 Reagents Kit (300 cycles), yielding a total of 2,081,668 raw paired-end reads with an average read length of 150 bp.

Raw read quality was inspected using FastP v0.23.4 (2). Genome assembly was completed with SPAdes v3.15.5 (3). The closest reference genome, *Bacillus safensis* GCF_008244765.1 (97.6%ANI), was identified by FastANI v1.34 (4) using reference genomes from the NCBI database (5). The resulting assembly was scaffolded with RagTag v2.1.0 (6) using the reference genome GCF_008244765.1. Contigs shorter than 500 bp were removed using SeqKit (7). The final draft genome assembly comprised 3,702,580 bp distributed across nine scaffolds with an average read coverage of 84.326x. Assembly metrics evaluated with QUAST v5.3.0 (8) reported N50 = 3,693,458 bp, L50 = 1, and a GC content of 41.67%. Genome completeness assessed with BUSCO v5.2.2 (9) with the bacillales_odb10 lineage data set indicated 99.8% complete single-copy BUSCOs.

Annotation of the assembled genome with the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) v6.10 (10) predicted 3,742 protein-coding genes. Secondary metabolite biosynthetic gene clusters were analyzed with antiSMASH v8.0 (11), revealing 14 clusters involved in various biosynthetic pathways. These included siderophore-associated clusters, such as those predicted to synthesize bacillibactin and schizokinen, which generate iron-chelating metabolites crucial for iron acquisition and often linked to antimicrobial activity (12). Additional clusters related to biosurfactant and antimicrobial

Editor Leighton Pritchard, University of Strathclyde, Glasgow, United Kingdom

Address correspondence to Vladimir K. Chebotar, vladchebotar@arriam.ru.

The authors declare no conflict of interest.

Received 18 November 2025

Accepted 26 March 2026

Published 20 April 2026

Copyright © 2026 Chebotar et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](#).

compound production, including lichenysin and fengycin, were also identified (13). Unless otherwise noted, default parameters were used for all software.

ACKNOWLEDGMENTS

This work was supported by the Russian Science Foundation (grant no. 23-66-10013) and performed using equipment of the Core Centrum "Genomic Technologies, Proteomics and Cell Biology" in ARRIAM. Sequencing was performed using the core facilities of the Lopukhin FRCC PCM "Genomics, Proteomics, Metabolomics" (<http://rcpcm.org/?p=2806>).

AUTHOR AFFILIATIONS

¹All-Russia Research Institute for Agricultural Microbiology, Pushkin, Russia

²Peter the Great St. Petersburg Polytechnic University, Saint Petersburg, Russia

³Saint Petersburg State University, Saint Petersburg, Russia

AUTHOR ORCIDs

Vladimir K. Chebotar  <http://orcid.org/0000-0001-9762-989X>

Maksim R. Losev  <http://orcid.org/0000-0003-0968-8684>

AUTHOR CONTRIBUTIONS

Vladimir K. Chebotar, Conceptualization, Funding acquisition, Project administration, Resources, Writing – original draft, Writing – review and editing | Veronika D. Antonova, Data curation, Investigation, Project administration | Maksim R. Losev, Formal analysis, Investigation, Methodology | Maria S. Gancheva, Conceptualization, Investigation, Methodology, Project administration, Writing – original draft, Writing – review and editing | Dmitriy V. Kudryavtsev, Data curation, Investigation, Resources | Olga A. Bortsova, Formal analysis, Resources, Validation | Elena P. Chizhevskaya, Data curation, Methodology, Resources, Visualization | Veronika N. Pishchik, Formal analysis, Project administration, Resources, Validation

DATA AVAILABILITY

This project has been deposited at GenBank under bioproject number [PRJNA1242717](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1242717). Whole genome assembly was deposited under accession number [GCF_054073805.1](https://www.ncbi.nlm.nih.gov/assembly/GCF_054073805.1). The raw reads were deposited in the Sequence Read Archive under accession number [SRR32881993](https://www.ncbi.nlm.nih.gov/sra/SRR32881993). antiSMASH annotation is available on Figshare (<https://doi.org/10.6084/m9.figshare.29038718>).

REFERENCES

1. Fu X, Gong L, Liu Y, Lai Q, Li G, Shao Z. 2021. *Bacillus pumilus* group comparative genomics: toward pangenome features, diversity, and marine environmental adaptation. *Front Microbiol* 12:571212. <https://doi.org/10.3389/fmicb.2021.571212>
2. Chen S. 2023. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *iMeta* 2:e107. <https://doi.org/10.1002/imt2.107>
3. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. 2020. Using SPAdes de novo assembler. *Curr Protoc Bioinformatics* 70:e102. <https://doi.org/10.1002/cpbi.102>
4. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* 9:5114. <https://doi.org/10.1038/s41467-018-07641-9>
5. Alonge M, Lebeigle L, Kirsche M, Jenike K, Ou S, Aganezov S, Wang X, Lippman ZB, Schatz MC, Soyk S. 2022. Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. *Genome Biol* 23:258. <https://doi.org/10.1186/s13059-022-02823-7>
6. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, et al. 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* 44:D733–45. <https://doi.org/10.1093/nar/gkv1189>
7. Shen W, Sipos B, Zhao L. 2024. SeqKit2: a Swiss army knife for sequence and alignment processing. *iMeta* 3:e191. <https://doi.org/10.1002/imt2.191>
8. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29:1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>
9. Manni M, Berkeley MR, Seppely M, Simão FA, Zdobnov EM. 2021. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol* 38:4647–4654. <https://doi.org/10.1093/molbev/msab199>
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>

11. Blin K, Shaw S, Augustijn HE, Reitz ZL, Biermann F, Alanjary M, Fetter A, Terlouw BR, Metcalf WW, Helfrich EJN, van Wezel GP, Medema MH, Weber T. 2023. antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic Acids Res* 51:W46–W50. <https://doi.org/10.1093/nar/gkad344>
12. Kramer J, Özkaya Ö, Kümmerli R. 2020. Bacterial siderophores in community and host interactions. *Nat Rev Microbiol* 18:152–163. <https://doi.org/10.1038/s41579-019-0284-4>
13. Satpute SK, Bhuyan SS, Pardesi KR, Mujumdar SS, Dhakephalkar PK, Shete AM, Chopade BA. 2010. Molecular genetics of biosurfactant synthesis in microorganisms. *Adv Exp Med Biol* 672:14–41. https://doi.org/10.1007/978-1-4419-5979-9_2