

Draft genome sequence of epiphyte *Paenibacillus kyungheensis* strain Cap286 isolated from *Chenopodium album* L. leaves

Vladimir K. Chebotar,¹ Maksim R. Losev,¹ Maria S. Gancheva,^{1,2} 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 *Paenibacillus kyungheensis* bacterium isolated from *Chenopodium album* L. leaves. The raw sequencing data were obtained using the NextSeq 1000 technology (Illumina). The assembly comprises two scaffolds, totaling 5,257,300 bp with a guanine-cytosine content of 39%.

KEYWORDS genome sequence, epiphyte, *Paenibacillus kyungheensis*, *Chenopodium album*, leaves

Paenibacillus kyungheensis strain Cap286, an epiphytic gram-positive bacterium, was isolated from the surface of the leaves of *Chenopodium album* collected in the Stavropol region of Russia (N_45°23'36", E_41°45'52") by placing plant leaf material (10 g) in 100 mL of sterile water and subjecting it to ultrasonic treatment (Bandelin; 50 Hz) for 10 min. Aliquots (0.1 mL) of serial dilutions were plated onto LB agar (Sigma-Aldrich, USA) and incubated for 3–5 days at 28°C.

Genomic DNA was extracted from single colony using the Wizard Genomic DNA Purification Kit (Promega, USA). A sequencing library was then prepared with the KAPA HyperPlus Kit (Roche, Switzerland). Purification was performed using KAPA Hyperpure beads (Roche, Switzerland). The quality and size distribution of the library were assessed with the High-Sensitivity DNA Kit (Agilent Technologies, USA), while its concentration was determined using the Quant-iT DNA Assay Kit (Thermo Fisher Scientific, USA). The sample was sequenced on an Illumina NextSeq 1000 instrument using a NextSeq 1000/2000 P1 Reagent Kit (300 cycles), which produced 2,075,020 paired-end reads with an average length of 150 bp.

The quality of the reads was checked using FastQC v0.11.9 (1). The draft genome was obtained with the SPAdes v4.2.0 genome assembler (2). FastANI v1.34 (3) identified that the closest reference genome was the genome of *Paenibacillus kyungheensis* GCF_028606985.1 (97.5% average nucleotide identity [ANI]) from the National Center for Biotechnology Information (NCBI) database (4). Scaffolding was performed using Ragtag v2.1.0 (5). Contigs shorter than 1 kbp were removed using the reformat tool from BBMap v39.08 (<https://sourceforge.net/projects/bbmap/>). Reads were mapped to the assembly using bwa v0.7.18 (6), and read coverage was calculated using the samtools coverage tool v1.13 (7). Assembly statistics were evaluated using QUAST v5.1.0 (8). The final assembly consisted of two scaffolds with a total length of 5,257,300 bp, guanine-cytosine content of 38.93%, N50 of 5,256,028, and average genome coverage of 64.8x. The completeness of genome was evaluated using BUSCO v5.2.2 (9) with the bacillales_odb10 lineage data set, and the BUSCO score was 99.8% (complete BUSCOs). At the same time, CheckM v1.2.3 (10) indicated 87.73% completeness and 1.7% contamination. A total of 4,683 genes were predicted on the assembly by NCBI Prokaryotic Genome Annotation Pipeline (PGAP) v6.10 (11). Unless otherwise noted, default parameters were used for all software.

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

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

The authors declare no conflict of interest.

Received 29 March 2026

Accepted 13 April 2026

Published 6 May 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](https://creativecommons.org/licenses/by/4.0/).

Secondary metabolite biosynthetic gene clusters (BGCs) were identified using antiSMASH v8.0 (12). The analysis revealed eight predicted clusters, including several for ribosomally synthesized and post-translationally modified peptides (RiPPs). Among these, a cluster for paeninodin was predicted. Paeninodin is a member of the lasso peptide class known for diverse bioactivities, such as antimicrobial, antiviral, and enzyme inhibitory functions (13). A separate cluster was predicted for the biosynthesis of a Class II lanthipeptide, mersacidin, which inhibits peptidoglycan synthesis (14). Additional clusters were associated with carotenoid and siderophore biosynthesis.

The significant role of Cap286 in improving wheat drought tolerance (15) underscores the importance of the genomic data presented here. This genome assembly provides a critical resource for future studies aimed at understanding, optimizing, and potentially engineering this strain for sustainable agriculture.

ACKNOWLEDGMENTS

This project was supported by the Russian Science Foundation Project No. 23-66-10013.

This research was 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, St. Petersburg, Pushkin, Russia

²Saint Peterburg State University, St. 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, Project administration, Resources, Supervision, Writing – review and editing | Maksim R. Losev, Formal analysis, Investigation, Methodology, Visualization | Maria S. Gancheva, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft | Dmitriy V. Kudryavtcev, Formal analysis, Validation, Visualization | Olga A. Bortsova, Data curation, Investigation, Validation | Elena P. Chizhevskaya, Conceptualization, Methodology, Validation, Visualization | Veronika N. Pishchik, Conceptualization, Data curation, Formal analysis, Investigation, Project administration, Validation, Visualization

DATA AVAILABILITY

The genome sequence and raw reads were deposited in NCBI and can be accessed under BioProject [PRJNA1337296](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1337296) (SRA [SRR35708438](https://www.ncbi.nlm.nih.gov/sra/SRR35708438), assembly [GCF_053570845.1](https://www.ncbi.nlm.nih.gov/genbank/GCF_053570845.1)).

REFERENCES

1. 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>
2. Pribelski 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>
3. 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>
4. O’Leary NA, Wright MW, Brister JR, Ciuffo S, Haddad D, McVeigh R, Rajput B, Robertse 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>
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. Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arxiv. <https://doi.org/10.48550/arXiv.1303.3997>
7. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H. 2021. Twelve years of SAMtools and BCFtools. *Gigascience* 10:giab008. <https://doi.org/10.1093/gigascience/giab008>
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. 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>
11. 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>
12. 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>
13. Zhu S, Hegemann JD, Fage CD, Zimmermann M, Xie X, Linne U, Marahiel MA. 2016. Insights into the unique phosphorylation of the lasso peptide Paeninodin. *Journal of Biological Chemistry* 291:13662–13678. <https://doi.org/10.1074/jbc.M116.722108>
14. Brötz H, Bierbaum G, Leopold K, Reynolds PE, Sahl HG. 1998. The lantibiotic mersacidin inhibits peptidoglycan synthesis by targeting lipid II. *Antimicrob Agents Chemother* 42:154–160. <https://doi.org/10.1128/AAC.42.1.154>
15. Pishchik VN, Chizhevskaya EP, Chebotar VK, Mirskaya GV, Khomyakov YV, Vertebny VE, Kononchuk PY, Kudryavtsev DV, Bortsova OA, Lapenko NG, Tikhonovich IA. 2024. PGPB isolated from drought-tolerant plants help wheat plants to overcome osmotic stress. *Plants (Basel)* 13:3381. <https://doi.org/10.3390/plants13233381>