

Complete genome sequence of *Planococcus koreensis* isolated from soil in Fort Collins, Colorado

Paige Gruber,¹ Ashley Freedman,¹ Kendall Malmstrom,¹ Bradley R. Borlee,¹ Carolina Mehaffy¹

AUTHOR AFFILIATION See affiliation list on p. 3.

ABSTRACT The complete genome of *Planococcus koreensis* was obtained using Nanopore MinION sequencing after isolation from soil in Colorado. The assembled genome contains one circular contig with 3,519,105 bp, 3,606 genes, 419 pseudogenes, and 47.62% guanine-cytosine content. This discovery provides a fully assembled *P. koreensis* genome available at the National Center for Biotechnology Information.

KEYWORDS Nanopore, *Planococcus koreensis*, genome

Whole-genome sequencing was performed as part of a Course-based Undergraduate Research Experience (CURE) at the Department of Microbiology, Immunology, and Pathology, Colorado State University. On 27 August 2024, an orange-pigmented bacteria isolate was obtained from a soil sample taken near Dixon Reservoir, Fort Collins, CO, at coordinates 40°33'12" N 105°8'25" W, at 5,200 ft elevation. After plating on tryptic soy agar for 24–48 h at room temperature, the isolate was streaked for isolation and identified as *Planococcus koreensis* via matrix-assisted laser desorption/ionization - time of flight (score = 1.74) (Bruker). *Planococcus* spp. produced a variety of metabolites with potential utility for bioremediation and industrial applications (1–3).

The isolate stained gram variable, as confirmed by previous literature (4). The isolate was grown in LB for 48 h before genomic DNA (gDNA) isolation using the Monarch Genomic DNA Purification Kit (#T3010S). Given the isolate's gram variable status, G+ and G– extraction protocols were performed. Library preparation from each gDNA preparation was performed using the Oxford Nanopore Technologies Rapid PCR Barcoding Kit 24 v.14 (SQK-RPB114.24) using the manufacturer's protocol and 25 PCR cycles. gDNA was not size-selected. Quality was checked via agarose gel electrophoresis and Qubit (dsDNA HS Assay Kit [Q32851]). Sequencing was performed with Oxford Nanopore Technologies MinION Flow Cell R10 Version (FLO-MIN114) using MinKnow v.24.06.8, Basecalling Fast Model v.4.3.0, 400 bp, and Min Q score of 8. Barcode trimming was performed post-run in MinKnow v.24.06.8 with default parameters.

Assembly and annotation were performed using Galaxy Server v.24.1.4.dev0 (5–7). Raw reads were pre-processed using Necat v.0.0.1_update20200803 + galaxy0 (8) in UseGalaxy.eu with default parameters and an estimated genome size of 3.5 Mb. All other tools were run on UseGalaxy.org. Concatenate Multiple Data Sets v.0.2 (9) was used to concatenate the corrected reads, followed by assembly with Flye v.2.9.5 + galaxy0 (10, 11), revealing a circular genome with one contig of 3,537,709 bp and coverage of 53%. Quast v.5.2.0 + galaxy1 (12–15) was run and indicated the contig had an N50 = 3,537,709. The assembly was polished using Racon v.1.5.0+galaxy1 (16), resulting in a polished genome of 3,519,105 bp. Annotation was performed using Prokka v.5.2.0 + galaxy1 (17, 18) with kingdom: bacteria. The isolate was confirmed as *Planococcus koreensis* by comparing the 16S rDNA gene from the annotated genome against the 16S rDNA bacteria and archaea database using BLAST (19) (99.87% identity and 97% query cover).

Editor Simon Roux, DOE Joint Genome Institute, Berkeley, California, USA

Address correspondence to Carolina Mehaffy, carolina.mehaffy@colostate.edu.

The authors declare no conflict of interest.

See the funding table on p. 3.

Received 21 February 2025

Accepted 15 March 2025

Published 17 April 2025

Copyright © 2025 Gruber 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/).

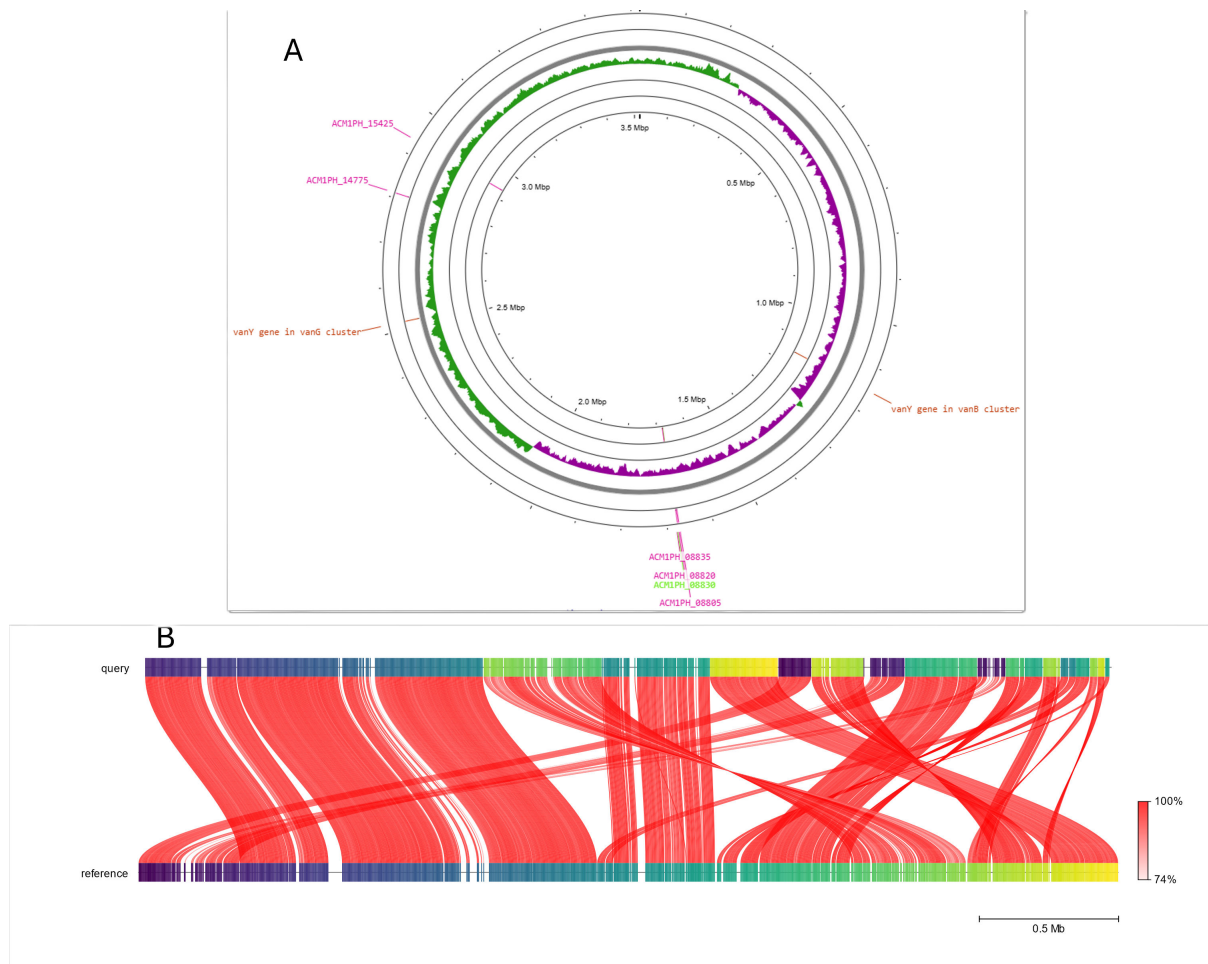


FIG 1 (A) *P. koreensis*, Fort Collins isolate, complete genome. Features show two predicted glycopeptide resistance genes (*vanY* genes), five genes, and one pseudogene predicted to be involved in carotenoid biosynthesis. Guanine-cytosine (GC) skew is shown (GC skew, purple; GC skew+, green). (B) Results from FastANI (average nucleotide identity [ANI]) against the *P. koreensis* reference genome (accession number JACHHE01) resulting in an ANI of 97.1598%. Each red line segment denotes a reciprocal mapping between the query (top) and reference (bottom) genomes. Shade of red represents the ANI percentage. Colors on the query and reference genome (horizontal segments) represent orthologous fragments of 3,000 bp.

The complete genome was visualized in Proksee (20) using the consensus fasta file (Fig. 1). FastANI v.1.1.0 (21) was used to compare the isolate genome with the reference, incomplete, 22-contig genome available at the National Center for Biotechnology Information (NCBI), *Planococcus koreensis* (accession number JACHHE01), resulting in 97.1598% average nucleotide identity with 1,061 orthologous matches out of 1,144 query sequence fragments (Fig. 1). The genome of *Planococcus koreensis* presented in this report is 3,519,105 bp, with a guanine-cytosine content of 47.62%, assembled in one circular contig (N50 = 3,519,105 bp). CARD analysis (<https://card.mcmaster.ca/analyze/rgi>) (22) identified two genes associated with glycopeptide resistance. Upon submission to the NCBI, the genome was annotated using the NCBI Prokaryotic Genome Annotation Pipeline (23–25), resulting in 3,606 coding sequences and 419 pseudogenes, including 5 genes and 1 pseudogene predicted to be carotenoid biosynthetic genes (Fig. 1).

ACKNOWLEDGMENTS

The authors thank the Department of Microbiology, Immunology, and Pathology at Colorado State University and the One Health Institute's BROADN project (NSF Biology Integration Institute #2120117) for funding and administrative support.

AUTHOR AFFILIATION

¹Department of Microbiology, Immunology and Pathology, Colorado State University, Fort Collins, Colorado, USA

AUTHOR ORCIDs

Bradley R. Borlee  <http://orcid.org/0000-0002-6321-2014>

Carolina Mehaffy  <http://orcid.org/0000-0002-7634-5131>

FUNDING

Funder	Grant(s)	Author(s)
National Science Foundation	2120117	Bradley R. Borlee

AUTHOR CONTRIBUTIONS

Paige Gruber, Formal analysis, Methodology, Writing – original draft, Writing – review and editing | Ashley Freedman, Methodology, Writing – review and editing | Kendall Malmstrom, Methodology, Writing – review and editing | Bradley R. Borlee, Investigation, Resources, Writing – review and editing.

DATA AVAILABILITY

This whole-genome assembly has been deposited at GenBank with accession number [ASM4779632v1](https://www.ncbi.nlm.nih.gov/assembly/ASM4779632v1) under biosample [SAMN46743435](https://www.ncbi.nlm.nih.gov/biosample/SAMN46743435). The annotated genome can be accessed under accession number [CP181055.1](https://www.ncbi.nlm.nih.gov/assembly/CP181055.1). The version described in this paper is the first version. Raw reads were submitted to the National Center for Biotechnology Information Sequence Read Archive with accession number [SRR32300276](https://www.ncbi.nlm.nih.gov/sra/SRR32300276).

REFERENCES

1. Ay H, Sahin SM, Duman M, Ajmi N, Saticioglu IB. 2025. Comprehensive genome analysis of sp. S3-L1 isolated from Horseshoe island in Antarctica reveals its biotechnological potential. *Turk J Fish Aquat Sci* 25. <https://doi.org/10.4194/TRJFAS27193>
2. Moyo AC, Dufossé L, Giuffrida D, van Zyl LJ, Trindade M. 2022. Structure and biosynthesis of carotenoids produced by a novel *Planococcus* sp. isolated from South Africa. *Microb Cell Fact* 21:43. <https://doi.org/10.1186/s12934-022-01752-1>
3. Waghmode S, Suryavanshi M, Sharma D, Satpute SK. 2020. *Planococcus* species - an imminent resource to explore biosurfactant and bioactive metabolites for industrial applications. *Front Bioeng Biotechnol* 8:996. <https://doi.org/10.3389/fbioe.2020.00996>
4. Hao S, Ren Q, Wang J, Li L, Huang M. 2023. Two novel *Planococcus* species isolated from Baijiu pit mud with potential application in brewing. *Front Microbiol* 14:1139810. <https://doi.org/10.3389/fmicb.2023.1139810>
5. Blankenberg D, Coraor N, Von Kuster G, Taylor J, Nekrutenko A, Galaxy Team. 2011. Integrating diverse databases into a unified analysis framework: a Galaxy approach. *Database (Oxford)* 2011:bar011. <https://doi.org/10.1093/database/bar011>
6. Galaxy C. 2024. The galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update. *Nucleic Acids Res* 52:W83–W94. <https://doi.org/10.1093/nar/gkae410>
7. Jalili V, Afgan E, Gu Q, Clements D, Blankenberg D, Goecks J, Taylor J, Nekrutenko A. 2020. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2020 update. *Nucleic Acids Res* 48:W395–W402. <https://doi.org/10.1093/nar/gkaa434>
8. Chen Y, Nie F, Xie SQ, Zheng YF, Dai Q, Bray T, Wang YX, Xing JF, Huang ZJ, Wang DP, He LJ, Luo F, Wang JX, Liu YZ, Xiao CL. 2021. Efficient assembly of nanopore reads via highly accurate and intact error correction. *Nat Commun* 12:60. <https://doi.org/10.1038/s41467-020-20236-7>
9. Goecks J, Nekrutenko A, Taylor J, Galaxy T. 2010. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol* 11:R86. <https://doi.org/10.1186/gb-2010-11-8-r86>
10. Kolmogorov M. 2021. Flye. GitHub. Available from: <https://github.com/fergusglass/Flye>
11. Lin Y, Yuan J, Kolmogorov M, Shen MW, Chaisson M, Pevzner PA. 2016. Assembly of long error-prone reads using de Bruijn graphs. *Proc Natl Acad Sci USA* 113:E8396–E8405. <https://doi.org/10.1073/pnas.1604560113>
12. 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>
13. Mikheenko A, Pribelski A, Saveliev V, Antipov D, Gurevich A. 2018. Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics* 34:i142–i150. <https://doi.org/10.1093/bioinformatics/bty266>
14. Mikheenko A, Saveliev V, Gurevich A. 2016. MetaQUAST: evaluation of metagenome assemblies. *Bioinformatics* 32:1088–1090. <https://doi.org/10.1093/bioinformatics/btv697>
15. Mikheenko A, Valin G, Pribelski A, Saveliev V, Gurevich A. 2016. Icarus: visualizer for de novo assembly evaluation. *Bioinformatics* 32:3321–3323. <https://doi.org/10.1093/bioinformatics/btw379>
16. Vaser R, Sović I, Nagarajan N, Šikić M. 2017. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res* 27:737–746. <https://doi.org/10.1101/gr.214270.116>
17. Cuccuru G, Orsini M, Pinna A, Sbardellati A, Soranzo N, Travaglione A, Uva P, Zanetti G, Fotia G. 2014. Orione, a web-based framework for NGS analysis in microbiology. *Bioinformatics* 30:1928–1929. <https://doi.org/10.1093/bioinformatics/btu135>
18. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>
19. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402. <https://doi.org/10.1093/nar/25.17.3389>

20. Grant JR, Enns E, Marinier E, Mandal A, Herman EK, Chen C, Graham M, Van Domselaar G, Stothard P. 2023. Proksee: in-depth characterization and visualization of bacterial genomes. *Nucleic Acids Res* 51:W484–W492. <https://doi.org/10.1093/nar/gkad326>
21. 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>
22. Alcock BP, Huynh W, Chalil R, Smith KW, Raphenya AR, Wlodarski MA, Edalatmand A, Petkau A, Syed SA, Tsang KK, et al. 2023. CARD 2023: expanded curation, support for machine learning, and resistome prediction at the comprehensive antibiotic resistance database. *Nucleic Acids Res* 51:D690–D699. <https://doi.org/10.1093/nar/gkac920>
23. Haft DH, DiCuccio M, Badretdin A, Brover V, Chetvernin V, O'Neill K, Li W, Chitsaz F, Derbyshire MK, Gonzales NR, Gwadz M, Lu F, Marchler GH, Song JS, Thanki N, Yamashita RA, Zheng C, Thibaud-Nissen F, Geer LY, Marchler-Bauer A, Pruitt KD. 2018. RefSeq: an update on prokaryotic genome annotation and curation. *Nucleic Acids Res* 46:D851–D860. <https://doi.org/10.1093/nar/gkx1068>
24. Li W, O'Neill KR, Haft DH, DiCuccio M, Chetvernin V, Badretdin A, Coulouris G, Chitsaz F, Derbyshire MK, Durkin AS, Gonzales NR, Gwadz M, Lanczycki CJ, Song JS, Thanki N, Wang J, Yamashita RA, Yang M, Zheng C, Marchler-Bauer A, Thibaud-Nissen F. 2021. RefSeq: expanding the prokaryotic genome annotation pipeline reach with protein family model curation. *Nucleic Acids Res* 49:D1020–D1028. <https://doi.org/10.1093/nar/gkaa1105>
25. 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>