ASM | Microbiology Resource Announcements

Check for updates

| Genomics and Proteomics | Announcement

Metabolic potential for benzoate utilization revealed by the complete genome of *Sphingobacterium* sp. PAMC25397

Min-Su Kim,[1] Byeollee Kim,[2] So-Ra Han,[2] Hyun Lee,[3] Hye Jin Jung,[4] Jun Hyuck Lee,[5] Tae-Jin Oh[1,2,4]

AUTHOR AFFILIATIONS See affiliation list on p. 3.

ABSTRACT Here, we report the complete genome sequence of *Sphingobacterium* sp. PAMC25397, isolated from Ny-Ålesund, Spitsbergen, Norway. The genome is 5.2 Mb in length with a GC content of 37%. This genome harbors biodegradation-related genes involved in benzoate degradation.

KEYWORDS *Sphingobacterium* sp., complete genome, benzoate degradation, Arctic bacteria

Aromatic compounds account for over 65% of mass-produced chemicals and organic pollutants (1). Benzoic acid is a common pollutant in dye effluents and food-processing wastewater (2). Microbial degradation of aromatics is important for waste treatment and bioremediation (3).

Sphingobacterium sp. PAMC25397 was isolated from animal excrement-associated debris collected in Ny-Ålesund, Spitsbergen, Norway (78.91502°N, 11.87143°E). Approximately 1 g of the sample was aseptically collected, cultured on pork-supplemented 0.1× nutrient agar at 4°C for 6 months, and purified by repeated streaking. The same single colony was used for both 16S rRNA gene amplification and whole-genome sequencing, and genomic DNA was extracted from R2A-grown cells using the DNeasy Blood & Tissue Kit (Qiagen, the Netherlands). The 16S rRNA gene was amplified using the universal primers 27F and 1492R and analyzed by Sanger sequencing (Cosmo Genetech, Korea). Taxonomic identification was based on comparison of the 16S rRNA sequence with the EzBioCloud database (4), and the closest hit was *Sphingobacterium* sp. ML3W (CP009278) (99.86%).

The same genomic DNA extract from PAMC25397 was used to prepare Illumina and Oxford Nanopore Technology (ONT) libraries for sequencing (Phyzen, Korea). Illumina libraries were prepared using the TruSeq DNA Nano Kit (Illumina, USA) and sequenced on a NovaSeq 6000 platform (2 × 150 bp paired-end reads). ONT libraries were prepared using the 1D Ligation Sequencing Kit (SQK-LSK110) and Native Barcoding Expansion Kit (EXP-NBD104) without size selection, sequenced on the GridION platform (FLO-MIN106, R9.4 flow cell), and basecalled with MinKNOW v22.12.7 and Guppy 6.4.6, both in high-accuracy mode. Pre-sequencing library QC was performed using Qubit and gel electrophoresis for ONT and qPCR and an Agilent 2100 Bioanalyzer (DNA 1000 chip) for Illumina. Illumina and Nanopore reads were processed using fastp v0.20.1 (default settings) and Filtlong v0.2.1 (–min_length 1000 and –keep_percent 95), respectively. ONT reads used for genome assembly comprised 53,309 reads, totaling 818.3 Mbp, with an N50 of 21,063 bp and approximately 159× coverage; Illumina reads comprised 8,644,901 reads, totaling 1,305.3 Mbp (Table 1).

Following the Trycycler workflow (5), 12 subsampled data sets were independently assembled using Flye v2.9.2 (6), reconciled using Trycycler v0.5.4 (7) to generate a consensus assembly, and then polished with Medaka v1.12.1 (8), Polypolish v0.6.0 (9), and POLCA v0.3.1 (10). Circularization was verified using Trycycler (11), which refines the

Editor Irene L. G. Newton, Indiana University Bloomington, Bloomington, Indiana, USA

Address correspondence to Tae-Jin Oh, tjoh3782@sunmoon.ac.kr.

The authors declare no conflict of interest.

See the funding table on p. 3.

Received 2 March 2026
Accepted 13 August 2026
Published 3 September 2026

Copyright © 2026 Kim et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

TABLE 1 Raw sequence information

Parameters	Genomic features
ONT statistics	
Reads	53,309
Total	818.3 Mbp
N50	21,063 bp
Coverage	159×
Illumina statistics	
Reads	8,644,901
Total	1,305.3 Mbp

contig boundaries. Genome quality was assessed using QUAST v5.2.0 (12) and CheckM v1.2.2 (13). CheckM estimated the genome to be 100% complete, with 1.43% contamination. Taxonomic classification was determined using the ANI Calculator (14). Genome annotation was performed using Prokka v1.14.6 (15) and eggnog-mapper v2.1.12 (16). KEGG pathway reconstruction was performed using KEGG Mapper (reconstruction module) (17). All software parameters were set to default unless otherwise specified.

The complete genome of PAMC25397 was assembled into a single circular contig of 5,157,497 bp and a GC content of 37%. ANI analysis revealed the highest similarity of 98.25% to *Sphingobacterium* sp. ML3W (GCA_000747525.1). Genome annotation identified a total of 4,387 protein-coding sequences, 89 tRNAs, and 1 transfer-messenger RNA (tmRNA. KEGG pathway reconstruction revealed multiple genes involved in benzoate degradation, suggesting genetic potential for this pathway (Fig. 1).

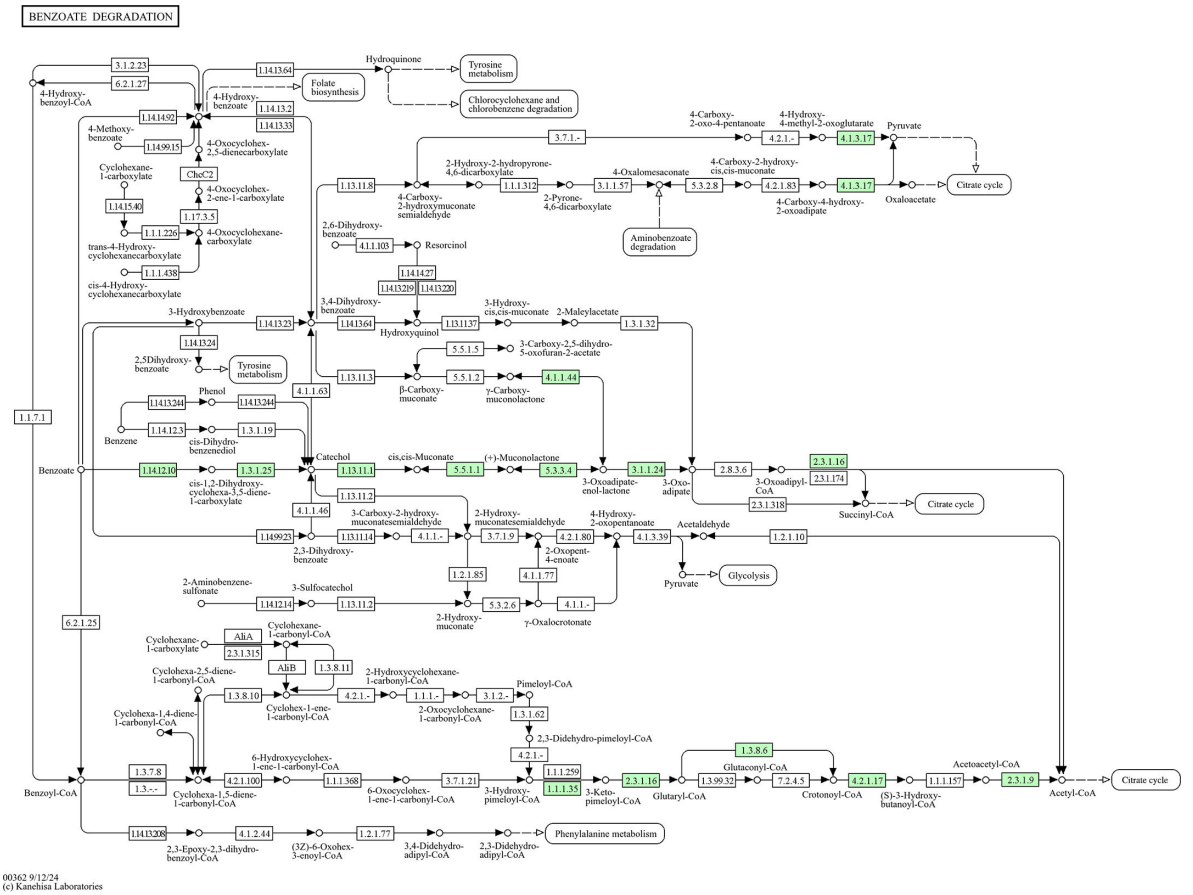


FIG 1 KEGG pathway reconstruction of benzoate degradation in *Sphingobacterium* sp. PAMC25397. Detected enzymes are highlighted in green.

ACKNOWLEDGMENTS

This research was supported by the Bio & Medical Technology Development Program of the National Research Foundation (NRF) funded by the Korean government (MSIT) (No. RS-2024-00441423). In addition, this work was supported by a Korea Polar Research Institute grant funded by the Ministry of Oceans and Fisheries (KOPRI Grant numbers PE26150 and PE26220). This work was also supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2025-24683148).

Min-Su Kim: Conceptualization, Data curation, Formal analysis, Investigation, Writing—original draft | Byeollee Kim: Data curation, Formal analysis, Validation | So-Ra Han: Data curation, Investigation, Validation | Hyun Lee: Data curation, Investigation, Validation | Hye Jin Jung: Data curation, Investigation, Validation | Jun Hyuck Lee: Data curation, Investigation, Validation | Tae-Jin Oh: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Validation, Writing—review and editing.

AUTHOR AFFILIATIONS

[1]Department of Life Science and Biochemical Engineering, Graduate School SunMoon University, Asan, South Korea
[2]Genome-based BioIT Convergence Institute, Asan, South Korea
[3]Division of Computer Science and Engineering, SunMoon University, Asan, South Korea
[4]Department of Pharmaceutical Engineering and Biotechnology, SunMoon University, Asan, South Korea
[5]Division of Life Sciences, Korea Polar Research Institute, Incheon, South Korea

AUTHOR ORCIDs

Jun Hyuck Lee http://orcid.org/0000-0002-4831-2228
Tae-Jin Oh http://orcid.org/0000-0002-8407-5039

FUNDING

Funder	Grant(s)	Author(s)
the Bio & Medical Technology Development Program of the National Research Foundation	RS-2024-00441423	Tae-Jin Oh
the Ministry of Oceans and Fisheries	PE26150	Jun Hyuck Lee

AUTHOR CONTRIBUTIONS

Min-Su Kim, Conceptualization, Data curation, Formal analysis, Investigation, Writing – original draft | Byeollee Kim, Data curation, Formal analysis, Validation | So-Ra Han, Data curation, Investigation, Validation | Hyun Lee, Data curation, Investigation, Validation | Hye Jin Jung, Data curation, Investigation, Validation | Jun Hyuck Lee, Data curation, Investigation, Validation | Tae-Jin Oh, Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Validation, Writing – review and editing

DATA AVAILABILITY

The complete genome sequencing for *Sphingobacterium* sp. PAMC25397 has been deposited in GenBank under the GCA_983196555.1. The raw sequencing reads for *Sphingobacterium* sp. PAMC25397 have been deposited in the ENA under the project accession number PRJEB106065, sample accession number SAMEA121008656, experiment accession number ERX15535198, and run accession number ERR16145640. The genome annotation files for PAMC25397, including Prokka output files and eggnog-mapper functional annotation files, have been deposited in Figshare and are available at https://doi.org/10.6084/m9.figshare.32165154.

REFERENCES

1. Howard PH. 1989. Handbook of environmental fate and exposure data for organic chemicals, volume 1—large production and priority pollutants, p 279–285. Lewis Publishers, Chelsea, MI.
2. Oie CSI, Albaugh CE, Peyton BM. 2007. Benzoate and salicylate degradation by *Halomonas campisalis*, an alkaliphilic and moderately halophilic microorganism. Water Res 41:1235–1242. https://doi.org/10.1016/j.watres.2006.12.029
3. Mekonnen BA, Aragaw TA, Genet MB. 2024. Bioremediation of petroleum hydrocarbon contaminated soil: a review on principles, degradation mechanisms, and advancements. Front Environ Sci 12:135422. https://doi.org/10.3389/fenvs.2024.1354422
4. Chalita M, Kim YO, Park S, Oh H-S, Cho JH, Moon J, Baek N, Moon C, Lee K, Yang J, Nam GG, Jung Y, Na S-I, Bailey MJ, Chun J. 2024. EzBioCloud: a genome-driven database and platform for microbiome identification and discovery. Int J Syst Evol Microbiol 74:006421. https://doi.org/10.1099/ijsem.0.006421
5. Wick RR, Judd LM, Holt KE. 2023. Assembling the perfect bacterial genome using Oxford Nanopore and illumina sequencing. PLoS Comput Biol 19:e1010905. https://doi.org/10.1371/journal.pcbi.1010905
6. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol 37:540–546. https://doi.org/10.1038/s41587-019-0072-8
7. Wick RR, Judd LM, Cerdeira LT, Hawkey J, Méric G, Vezina B, Wyres KL, Holt KE. 2021. Trycycler: consensus long-read assemblies for bacterial genomes. Genome Biol 22:266. https://doi.org/10.1186/s13059-021-02483-z
8. Wright C, Wykes M. 2023. Medaka: sequence correction provided by ONT research. https://github.com/nanoporetech/medaka.
9. Wick RR, Holt KE. 2022. Polypolish: short-read polishing of long-read bacterial genome assemblies. PLoS Comput Biol 18:e1009802. https://doi.org/10.1371/journal.pcbi.1009802
10. Zimin AV, Salzberg SL. 2020. The genome polishing tool POLCA makes fast and accurate corrections in genome assemblies. PLoS Comput Biol 16:e1007981. https://doi.org/10.1371/journal.pcbi.1007981
11. Crisan CV, Duncan RP, Van Tyne D, Goldberg JB. 2024. Complete genome sequence of *Stenotrophomonas maltophilia* complex strain STEN00241. Microbiol Resour Announc 13:e01053-23. https://doi.org/10.1128/mra.01053-23
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. 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
14. Yoon SH, Ha SM, Lim J, Kwon S, Chun J. 2017. A large-scale evaluation of algorithms to calculate average nucleotide identity. Antonie Van Leeuwenhoek 110:1281–1286. https://doi.org/10.1007/s10482-017-0844-4
15. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153
16. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. 2021. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Mol Biol Evol 38:5825–5829. https://doi.org/10.1093/molbev/msab293
17. Kanehisa M, Sato Y. 2020. KEGG mapper for inferring cellular functions from protein sequences. Protein Sci 29:28–35. https://doi.org/10.1002/pro.3711