

Draft genome sequence of *Muricoccus roseus* DSM 14916^T, isolated from a children's care center in Finland

Alejandra Valenzuela,¹ Jair Cortez,¹ Janelle Mande,¹ Cody Fondse,¹ Vera Thiel,² Kurt LaButti,³ Nikos Kyrpides,³ Rekha Seshadri,³ Tricia A. Van Laar¹

AUTHOR AFFILIATIONS See affiliation list on p. 3.

ABSTRACT *Muricoccus roseus* DSM 14916^T, a strain also reported as *Roseomonas rosea* and *Pararoseomonas rosea*, was isolated from a water-damaged childcare center in Finland. We report the whole-genome sequence and annotation, identifying genes associated with carotenoid biosynthesis. This sequence may be useful for studies on taxonomic clarification and environmental microbiology.

KEYWORDS *Muricoccus*, taxonomy, phylogeny, *Roseomonas*, *Pararoseomonas*

Muricoccus roseus DSM 14916^T, a pink-pigmented, Gram-negative bacterium, was isolated from a water-damaged wall of a childcare center in Finland (1) and classified as a novel genus and species based on 16S rRNA sequencing and physiological characteristics (2). Subsequent proposals placed this species in *Roseomonas* (3) and *Pararoseomonas* (4), but those reclassifications, published under the International Code of Nomenclature of Prokaryotes (5), are considered illegitimate by the List of Prokaryotic names with Standing in Nomenclature (6, 7). This inconsistency is highlighted by polyphyly of *Roseomonas* in Bergey's Manual (8). The genome of *M. roseus* DSM 14916^T was sequenced as part of the Genomic Encyclopedia of Bacteria and Archaea project to broaden genomic resource diversity (9).

DSM 14916^T was cultured aerobically for 4 days in TSB at 28°C and 100 rpm at the DSMZ. Biomass was harvested by centrifugation, and DNA was extracted using the MasterPure DNA Purification Kit (Lucigen) per the manufacturer's instructions. One hundred nanograms of DNA was sheared to 270 bp using Covaris LE220 and size-selected with SPRI beads (Beckman Coulter). Fragments underwent end-repair, A-tailing, and adapter ligation (IDT, Inc.) using KAPA-Illumina Library Creation Kit (KAPA Biosystems). The library was quantified using KAPA Library Quantification Kit (Roche) on Roche LightCycler. The library was then multiplexed with others and prepared for sequencing on the Illumina HiSeq sequencing platform utilizing a TruSeq Paired-end Cluster Kit, v3, and Illumina's cBot instrument to generate a clustered flow cell. Sequencing was performed on an Illumina HiSeq 2000 using HiSeq TruSeq SBS sequencing kits, v3, with a 2 × 150 indexed run recipe. DUK, part of BBtools, was run with default parameters for quality, adapter, and contaminant trimming (10). Reads were assembled using Velvet (v.1.2.07; velvet: 63 -shortPaired and velvetg: -very clean yes -exportFiltered yes -min contig lgth 500 -scaffolding no -cov cutoff 10) (11). Simulated 1–3 kb paired-end reads were created with wgsim (-e 0 -1 100 -2 100 -r 0 -R 0 -X 0) (12). Simulated and Illumina reads were assembled using Allpaths-LG (v.r46652; PrepareAllpathsInputs: PHRED 64=0 PLOIDY=1 FRAG COVERAGE=125 JUMP COVERAGE=25 LONG JUMP COV=50, RunAllpathsLG: THREADS=8 RUN=std shredpairs TARGETS=standard VAPI WARN ONLY=True OVERWRITE=True). The genome was 96.89% complete with 0.16% contamination via CheckM2 v.1.0.2 (13). Genome annotation used the JGI Microbial Genome Annotation

Editor John J. Dennehy, Queens College
Department of Biology, Queens, New York, USA

Address correspondence to Rekha Seshadri,
rseshadri@ibl.gov, or Tricia A. Van Laar,
tvanlaar@csustan.edu.

Alejandra Valenzuela, Jair Cortez, Janelle Mande, and
Cody Fondse contributed equally to this article.
Author order was determined by a random number
generator.

The authors declare no conflict of interest.

Received 8 August 2025

Accepted 29 October 2025

Published 26 November 2025

Copyright © 2025 Valenzuela et al. This is an
open-access article distributed under the terms of
the [Creative Commons Attribution 4.0 International
license](#).

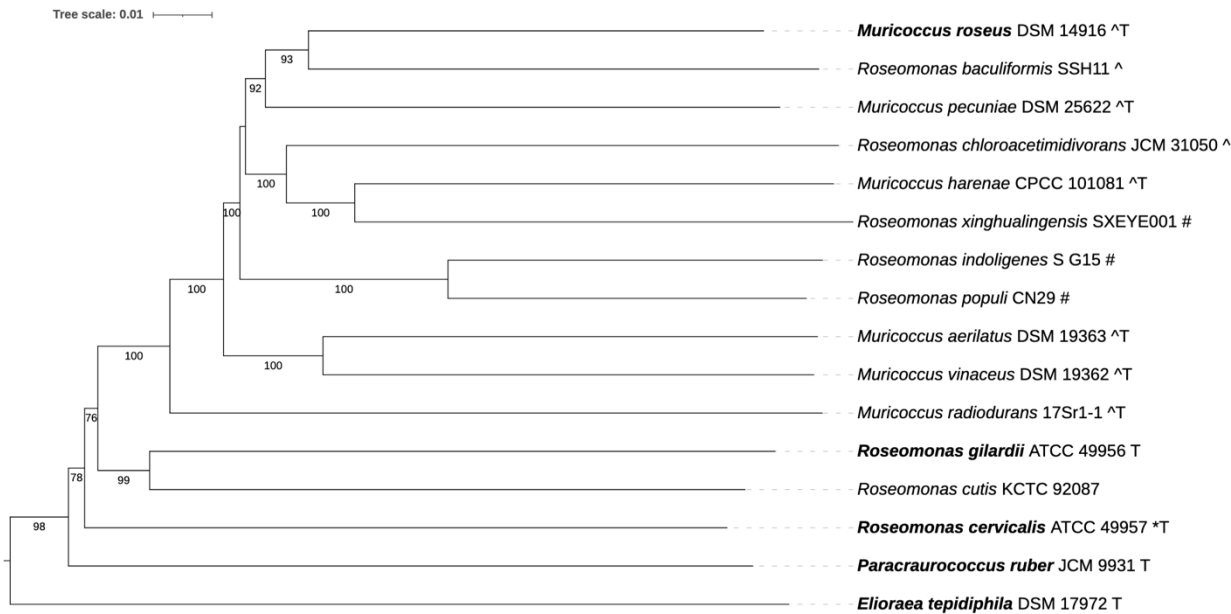


FIG 1 Tree inferred with FastME 2.1.6.1 (19) from Genome BLAST Distance Phylogeny (GBDP) distances calculated from genome sequences. The branch lengths are scaled using GBDP distance formula d_s . Numbers are GBDP pseudo-bootstrap support values >60% from 100 replications, with an average branch support of 92.5%. The tree was rooted using *Elioraea tepidiphila* as an outgroup (20). Bold, type species for genus; T, type strain for species; *, recommended reclassification to *Pseudoroseomonas* (4); ^, recommended reclassification to *Pararoseomonas* (4); #, likely *Pararoseomonas* but sequenced after Rai et al. (4).

TABLE 1 Genomic features of *Muricoccus roseus* DSM 14916^T

Feature	Finding
Genome size (bp)	5,342,582
Number of scaffolds	81
Scaffold N50 (bp)	151,080
Scaffold L50 (kb)	12
GC content (%)	70.76
Genome coverage	210.8×
No. of protein coding genes	5,027
tRNA genes	47
rRNA genes (5S, 16S, and 23S)	7 (1, 1, 5)

Pipeline v.4 (14) with the standard DOE-JGI protocol. Results were integrated into the Integrated Microbial Genomes system for comparative analysis (15).

The final genome assembly was 5,342,582 bp with a G+C content of 70.76% and 5,027 predicted protein-coding genes (Table 1). A phylogenetic tree was generated using the Type Strain Genome Server (16) and visualized using the Interactive Tree of Life v.6 (17) (Fig. 1). The closest relative of *M. roseus* was *R. baculiformis* (44.5% dDDH), with 17.3% similarity to *R. gilardii*, supporting its placement outside the genus *Roseomonas*. Given this isolate's pigmentation, we used antiSMASH (v.8.0) (18) which detected carotenoid biosynthesis genes, including *crtB* (*BUB97_RS25410*), *crtI* (*BUB97_RS02570*), and *crtX* (*BUB97_RS02555*). Additional clusters included a hydrogen cyanide operon (*hcnABC*: *BUB97_RS10025*, *BUB97_RS10030*, *BUB97_RS10035*) and a non-ribosomal peptide synthetase (*BUB97_RS12185*), suggesting potential for antimicrobial and bioactive compound production.

ACKNOWLEDGMENTS

The work (proposal DOI [10.46936/10.25585/60000886](https://doi.org/10.46936/10.25585/60000886)) conducted by the U.S. Department of Energy Joint Genome Institute (<https://ror.org/04xm1d337>), a DOE Office of

Science User Facility, is supported by the Office of Science of the U.S. Department of Energy operated under contract no. DE-AC02-05CH11231.

We thank JGI for their support of undergraduate education and research experiences at the California State University system through their “Adopt-a-Genome” project, allowing undergraduate students to take the lead on manuscript preparation.

AUTHOR AFFILIATIONS

¹Department of Biological Sciences, California State University, Stanislaus, Turlock, California, USA

²Leibniz Institute DSMZ—German Collection of Microorganisms and Cell Cultures GmbH, Braunschweig, Germany

³DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, California, USA

AUTHOR ORCIDs

Kurt LaButti  <http://orcid.org/0000-0002-5838-1972>

Nikos Kyrpides  <http://orcid.org/0000-0002-6131-0462>

Rekha Seshadri  <http://orcid.org/0000-0003-3219-2900>

Tricia A. Van Laar  <http://orcid.org/0000-0002-4654-8501>

AUTHOR CONTRIBUTIONS

Alejandra Valenzuela, Data curation, Formal analysis, Investigation, Software, Validation, Visualization, Writing – original draft | Jair Cortez, Data curation, Formal analysis, Investigation, Software, Validation, Visualization, Writing – original draft | Janelle Mande, Data curation, Formal analysis, Investigation, Software, Validation, Visualization, Writing – original draft | Cody Fondse, Data curation, Formal analysis, Investigation, Software, Validation, Visualization, Writing – original draft | Vera Thiel, Data curation, Methodology, Resources, Writing – review and editing | Kurt LaButti, Data curation, Methodology, Resources, Software, Writing – review and editing | Nikos Kyrpides, Data curation, Methodology, Resources, Software, Writing – review and editing | Rekha Seshadri, Conceptualization, Data curation, Funding acquisition, Project administration, Resources, Writing – review and editing | Tricia A. Van Laar, Conceptualization, Data curation, Project administration, Supervision, Validation, Visualization, Writing – original draft, Writing – review and editing

DATA AVAILABILITY

This Whole Genome project has been deposited in GenBank under accession no. [FQZF00000000](https://www.ncbi.nlm.nih.gov/nuclseq/FQZF01000000). The version described in this paper is the first version, [FQZF01000000](https://www.ncbi.nlm.nih.gov/nuclseq/FQZF01000000). The raw reads have been deposited in the NCBI Sequence Read Archive (SRA) under the accession number [SRX2066078](https://www.ncbi.nlm.nih.gov/sra/SRX2066078). Additional data can be explored or downloaded from the JGI Integrated Microbial Genomes with Microbiomes (IMG/M) portal using the taxon ID [2585427600](https://www.ncbi.nlm.nih.gov/taxonomy/2585427600).

REFERENCES

- Andersson MA, Nikulin M, Kõljalg U, Andersson MC, Rainey F, Reijula K, Hintikka EL, Salkinoja-Salonen M. 1997. Bacteria, molds, and toxins in water-damaged building materials. *Appl Environ Microbiol* 63:387–393. <https://doi.org/10.1128/aem.63.2.387-393.1997>
- Kämpfer P, Andersson MA, Jäckel U, Salkinoja-Salonen M. 2003. *Teichococcus ludpueritiae* gen. nov. sp. nov., and *Muricoccus roseus* gen. nov. sp. nov. representing two new genera of the α -1 subclass of the proteobacteria. *Syst Appl Microbiol* 26:23–29. <https://doi.org/10.1078/072320203322337272>
- Sánchez-Porro C, Gallego V, Busse H-J, Kämpfer P, Ventosa A. 2009. Transfer of *Teichococcus ludpueritiae* and *Muricoccus roseus* to the genus *Roseomonas*, as *Roseomonas ludpueritiae* comb. nov. and *Roseomonas rosea* comb. nov., respectively, and emended description of the genus *Roseomonas*. *Int J Syst Evol Microbiol* 59:1193–1198. <https://doi.org/10.1099/ijs.0.004820-0>
- Rai A, Jagadeeshwari U, Deepshikha G, Smita N, Sasikala C, Ramana CV. 2021. Phylotaxogenomics for the reappraisal of the genus *Roseomonas* with the creation of six new genera. *Front Microbiol* 12:677842. <https://doi.org/10.3389/fmicb.2021.677842>
- Oren A, Arahal DR, Göker M, Moore ERB, Rossello-Mora R, Sutcliffe IC. 2023. International code of nomenclature of prokaryotes. prokaryotic code (2022 Revision). *Int J Syst Evol Microbiol* 73. <https://doi.org/10.1099/ijsem.0.005585>

6. Meier-Kolthoff JP, Carbasse JS, Peinado-Olarte RL, Göker M. 2022. TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. *Nucleic Acids Res* 50:D801–D807. <https://doi.org/10.1093/nar/gkab902>
7. Parte AC, Sardà Carbasse J, Meier-Kolthoff JP, Reimer LC, Göker M. 2020. List of prokaryotic names with standing in nomenclature (LPSN) moves to the DSMZ. *Int J Syst Evol Microbiol* 70:5607–5612. <https://doi.org/10.1099/ijsem.0.004332>
8. Weyant RS, Whitney AM. 2015. *Roseomonas*, p 1–9. In *Bergey's manual of systematics of archaea and bacteria*
9. Kyrpides NC, Woyke T, Eisen JA, Garrity G, Lilburn TG, Beck BJ, Whitman WB, Hugenholtz P, Klenk H-P. 2014. Genomic encyclopedia of type strains, phase I: the one thousand microbial genomes (KMG-I) project. *Stand Genomic Sci* 9:1278–1284. <https://doi.org/10.4056/sigs.5068949>
10. Mingkun L, Copeland A, Han J. 2011. DUK – a fast and efficient Kmer based sequence matching tool. <https://www.osti.gov/servlets/purl/1016000>
11. Zerbino DR. 2010. Using the velvet *de novo* assembler for short - read sequencing technologies. *CP in Bioinformatics* 31:11. <https://doi.org/10.1002/0471250953.bi1105s31>
12. Li H. 2010. WgSim: a small tool for simulating sequence reads from a reference genome. Available from: <https://github.com/lh3/wgsim>
13. 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>
14. Huntemann M, Ivanova NN, Mavromatis K, Tripp HJ, Paez-Espino D, Palaniappan K, Szeto E, Pillay M, Chen I-M, Pati A, Nielsen T, Markowitz VM, Kyrpides NC. 2015. The standard operating procedure of the DOE-JGI microbial genome annotation pipeline (MGAP v.4). *Stand in Genomic Sci* 10:86. <https://doi.org/10.1186/s40793-015-0077-y>
15. Chen I-M, Chu K, Palaniappan K, Ratner A, Huang J, Huntemann M, Hajek P, Ritter SJ, Webb C, Wu D, Varghese NJ, Reddy TBK, Mukherjee S, Ovchinnikova G, Nolan M, Seshadri R, Roux S, Visel A, Woyke T, Elae-Fadros EA, Kyrpides NC, Ivanova NN. 2023. The IMG/M data management and analysis system v.7: content updates and new features. *Nucleic Acids Res* 51:D723–D732. <https://doi.org/10.1093/nar/gkac976>
16. 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>
17. Letunic I, Bork P. 2024. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res* 52:W78–W82. <https://doi.org/10.1093/nar/gkac268>
18. 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>
19. Lefort V, Desper R, Gascuel O. 2015. FastME 2.0: a comprehensive, accurate, and fast distance-based phylogeny inference program. *Mol Biol Evol* 32:2798–2800. <https://doi.org/10.1093/molbev/msv150>
20. Farris JS. 1972. Estimating phylogenetic trees from distance matrices. *Am Nat* 106:645–668. <https://doi.org/10.1086/282802>