

Lawrence Berkeley National Laboratory

LBL Publications

Title

Catalog of metagenome-assembled genomes of prokaryotic communities from the Red Sea hydrothermal vents

Permalink

<https://escholarship.org/uc/item/9rq516zq>

Journal

Microbiology Resource Announcements, 15(9)

ISSN

2169-8287

Authors

Schultz, Júnia

Altalhi, Sharifah

Camargo, Antônio Pedro

et al.

Publication Date

2026-09-10

DOI

10.1128/mra.00482-26

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed



Catalog of metagenome-assembled genomes of prokaryotic communities from the Red Sea hydrothermal vents

Júnia Schultz,^{1,2} Sharifah Altalhi,^{1,3} Antônio Pedro Camargo,⁴ Nikos C. Kyrpides,⁴ Alexandre Soares Rosado¹

AUTHOR AFFILIATIONS See affiliation list on p. 2.

ABSTRACT This study presents medium- and high-quality prokaryotic metagenome-assembled genomes (MAGs) from microbial mats and sediments at Hatiba Mons, a Red Sea hydrothermal system. We recovered 1,217 bacterial and archaeal MAGs across 75 phyla, dominated by *Planctomycetota* and *Thermoproteota*. Approximately 70% of these genomes likely represent previously uncharacterized taxa.

KEYWORDS metagenomics, microbial diversity, hydrothermal system, microbial genomes, extreme environment

The exploration of hydrothermal vent ecosystems has revealed insights into the diversity of life in extreme environments (1, 2). However, Red Sea hydrothermal systems remain underrepresented in genome-resolved studies. The recent discovery of the first active hydrothermal vent fields in the Red Sea offers new opportunities to elucidate the composition and functional potential of microbial communities inhabiting these systems (3, 4). Here, we report 1,217 medium- and high-quality prokaryotic metagenome-assembled genomes (MAGs) recovered from Hatiba Mons, Red Sea rift.

A total of 31 samples (19 sediments and 12 microbial mats) were collected from 5 hydrothermal vents (22.28°N, 37.84°E) during expeditions in May and September 2022 and February 2023. Sediment samples were collected using gravity corers, while microbial mats were sampled using push cores, Niskin bottles, and box corers, in triplicate. All samples were immediately frozen at −20°C onboard and subsequently stored at −80°C in the laboratory until processing. Further details are available in Altalhi et al. (4). DNA was extracted from 10 g of sample using the DNeasy PowerMax Soil Kit (Qiagen, Germany), following the manufacturer's protocol, with the addition of proteinase K (40 µL of 20 mg/mL, Omega) during lysis phase. DNA concentration was measured using the Qubit dsDNA HS Assay on a Qubit 4.0 fluorometer (Thermo Fisher Scientific, USA). For shotgun sequencing, extracted DNA ranging from 80 to 895 ng per sample was submitted to Novogene Inc. (California, USA) that prepared the library using the NEBNext Ultra II DNA Library Prep Kit (New England Biolabs, USA). Library preparation included DNA fragmentation, end repair, 5' phosphorylation, A-tailing, adapter ligation, and PCR amplification. The resulting libraries were sequenced on an Illumina NovaSeq platform, generating paired-end 150-bp reads.

Metagenomic sequencing generated a total of 3,004,009,760 paired-end reads across all samples (ranging from 76,449,854 to 303,487,370 reads per sample). Low-quality reads and adapters were removed using fastp v1.3.0 (5). Quality-filtered reads were assessed with FastQC v0.12.1 and assembled using MEGAHIT v1.2.9 (6), resulting in 31 individual assemblies and 11 co-assemblies. Contigs ≥ 1,000 bp were binned using MetaBAT2 v1.7 (7), MaxBin2 v2.2.7 (8), and VAMB v5.0.2 (9), followed by bin aggregation with DAS Tool v1.1.7 (10). Genome quality was assessed using CheckM2 v1.1.0 (11), and only high-quality and medium-quality MAGs (12) were retained for downstream analyses. These bins underwent taxonomic assignment via GTDB-Tk v2.4.1 (release 226) (13). Protein-coding

Editor Frank J. Stewart, Montana State University, Bozeman, Montana, USA

Address correspondence to Alexandre Soares Rosado, alexandre.rosado@kaust.edu.sa, or Júnia Schultz, juniaschultz@usp.br.

Júnia Schultz and Sharifah Altalhi contributed equally to this article. Author order was determined by seniority.

The authors declare no conflict of interest.

Received 17 April 2026

Accepted 17 July 2026

Published 4 August 2026

Copyright © 2026 Schultz 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/).

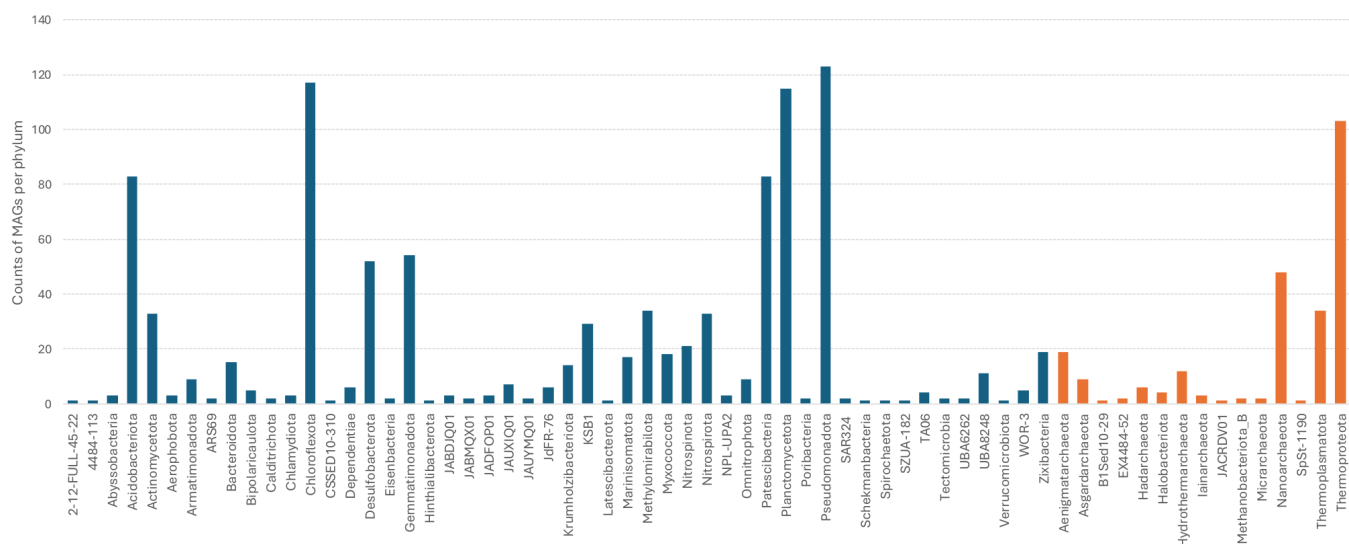


FIG 1 Distribution of metagenome-assembled genomes (MAGs) across prokaryotic phyla. The bar plot shows the number of recovered MAGs assigned to each bacterial (blue) and archaeal (orange) phylum based on GTDB taxonomy.

sequences were predicted from all MAGs using Prodigal (14). Default parameters were applied for all software.

A total of 1,217 dereplicated prokaryotic MAGs were recovered from microbial mats and sediments. Of these, 969 bacterial MAGs spanned 50 phyla, with *Pseudomonadota*, *Chloroflexota*, and *Planctomycetota* as the most abundant groups. Additionally, 248 archaeal MAGs were identified, distributed across 15 phyla, with *Thermoproteota*, *Thermoplasmatota*, and *Nanoarchaeota* as the dominant taxa (Fig. 1). Based on GTDB RED (relative evolutionary divergence), approximately 70% of the MAGs likely represent novel taxa, a proportion that exceeds or matches reports from other hydrothermal vent systems. These findings highlight the Hatiba Mons as a hotspot of microbial novelty and expand available genomic references.

ACKNOWLEDGMENTS

This work was supported by Prof. Alexandre Soares Rosado's KAUST Baseline Grant (BAS/1/1096-01-01). Júnia Schultz is grateful to the International Society for Microbial Ecology for partially funding her traineeship at the DOE-JGI, as well as to the National Council for Scientific and Technological Development (CNPq) (Grant numbers 314895/2025-6 and 446300/2024-1). The work conducted by Antônio Pedro Camargo and Nikos C. Kyrpides in the U.S. Department of Energy Joint Genome Institute (<https://ror.org/04xm1d337>) was supported by the U.S. Department of Energy Office of Science user facilities, operated under contract number DE-AC02-05CH11231.

AUTHOR AFFILIATIONS

¹Biological and Environmental Sciences and Engineering Division, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

²Center for Nuclear Energy in Agriculture, University of São Paulo, Piracicaba, Brazil

³Taif University, Taif, Saudi Arabia

⁴DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, California, USA

AUTHOR ORCIDs

Júnia Schultz  <http://orcid.org/0000-0002-8147-4659>

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

Alexandre Soares Rosado  <http://orcid.org/0000-0001-5135-1394>

AUTHOR CONTRIBUTIONS

Júnia Schultz, Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Validation, Visualization, Writing – original draft, Writing – review and editing | Sharifah Altalhi, Formal analysis, Methodology, Visualization, Writing – review and editing | Antônio Pedro Camargo, Formal analysis, Investigation, Methodology, Project administration, Resources, Supervision, Writing – review and editing | Nikos C. Kyrpides, Funding acquisition, Project administration, Resources, Supervision, Writing – review and editing | Alexandre Soares Rosado, Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – review and editing

DATA AVAILABILITY

The metagenomic data were deposited in the National Center for Biotechnology Information database under BioProject [PRJNA1121395](https://ncbi.nlm.nih.gov/bioproject/PRJNA1121395). BioSample accession numbers for the raw reads are from [SRX24841039](https://ncbi.nlm.nih.gov/biosample/SRX24841039) to [SRX24841069](https://ncbi.nlm.nih.gov/biosample/SRX24841069). MAG sequences and supplementary tables (metadata, quality, and taxonomy) are available in the open repository Zenodo (<https://doi.org/10.5281/zenodo.19235833>).

REFERENCES

- Dick GJ. 2019. The microbiomes of deep-sea hydrothermal vents: distributed globally, shaped locally. *Nat Rev Microbiol* 17:271–283. <https://doi.org/10.1038/s41579-019-0160-2>
- Schultz J, Altalhi S, Zwan F van der, Augustin N, Rosado AS. 2025. Diving into the deep: fungal diversity in the newly discovered hydrothermal vents of Hatiba Mons, Red Sea. *Front Mar Sci* 12:1649339. <https://doi.org/10.3389/fmars.2025.1649339>
- van der Zwan FM, Augustin N, Petersen S, Altalhi SM, Schultz J, Peixoto RS, Follmann J, Anker A, Benzoni F, Garcia Paredes ER, Al Malallah M, Shepard L, Ouhssain M, Jägerup SB, Jones BH, Rosado AS. 2023. Widespread diffuse venting and large microbial iron-mounds in the Red Sea. *Commun Earth Environ* 4. <https://doi.org/10.1038/s43247-023-0116-9>
- Altalhi S, Schultz J, Jamil T, Diercks I, Sharma S, Follmann J, Alam I, Raman K, Augustin N, van der Zwan FM, Rosado AS. 2025. Decoding microbial diversity, biogeochemical functions, and interaction potentials in red sea hydrothermal vents. *Environ Microbiome* 20:118. <https://doi.org/10.1186/s40793-025-00784-5>
- Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34:i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>
- Li D, Liu CM, Luo R, Sadakane K, Lam TW. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31:1674–1676. <https://doi.org/10.1093/bioinformatics/btv033>
- Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, Wang Z. 2019. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* 7:e7359. <https://doi.org/10.7717/peerj.7359>
- Wu YW, Simmons BA, Singer SW. 2016. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* 32:605–607. <https://doi.org/10.1093/bioinformatics/btv638>
- Nissen JN, Johansen J, Allesøe RL, Sønderby CK, Armenteros JJA, Grønbech CH, Jensen LJ, Nielsen HB, Petersen TN, Winther O, Rasmussen S. 2021. Improved metagenome binning and assembly using deep variational autoencoders. *Nat Biotechnol* 39:555–560. <https://doi.org/10.1038/s41587-020-00777-4>
- Siebert CMK, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, Banfield JF. 2018. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat Microbiol* 3:836–843. <https://doi.org/10.1038/s41564-018-0171-1>
- 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* 21:1203–1212. <https://doi.org/10.1038/s41592-024-02248-z>
- Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, Schulz F, Jarett J, Rivers AR, Elie-Fadrosh EA, et al. 2017. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol* 35:725–731. <https://doi.org/10.1038/nbt.3893>
- Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. 2022. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics* 38:5315–5316. <https://doi.org/10.1093/bioinformatics/bt ac672>
- Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. <https://doi.org/10.1186/1471-2105-11-119>