

Draft genome sequence of *Bacillus* sp. strain X and *Salarachaeum* sp. strain III isolated from Lake Karum, Danakil Depression, Ethiopia

Ermias Sissay Balcha,^{1,2} Karen Olsson-Francis,³ Ben Stephens,³ Barbara Cavalazzi,^{4,5} Aduugna Abdi Woldesemayat,² Mesfin Tafesse Gameda,² Michael C. Macey³

AUTHOR AFFILIATIONS See affiliation list on p. 2.

ABSTRACT Here, we report the draft genome sequences of strains of *Bacillus* and *Salarachaeum* that were isolated from hypersaline water samples collected from Lake Karum, Danakil Depression, Ethiopia. The sequences pave the way for more targeted studies into the potential biological activities and secondary metabolite synthesis of these organisms.

KEYWORDS genome sequences, extreme environments, Lake Karum

Lake Karum, also known as Lake Assale, is located in the Danakil Depression, northern Ethiopia. It is a brine lake (28%–33% salinity) that lies 120 m below sea level with temperatures ranging between 34°C and 36°C (1, 2). Previous studies have shown that Lake Karum supports a diverse microbial community (3, 4). Here, we present the draft genomes of *Bacillus* sp. strain X and *Salarachaeum* sp. strain III. The strains were isolated from water samples collected using falcon tubes from a depth of 5 cm in Lake Karum (14.0207° N, 40.4007° E).

The water samples were enriched with Soya Flour Medium (SFM) containing 4 M NaCl. Serial dilutions were prepared and plated onto soy flour agar and incubated at 30°C for 30 days. Distinct colonies were observed, leading to isolation of *Bacillus* sp. strain X and *Salarachaeum* sp. strain III (5). Pure colonies were used to culture in SFM, from which DNA was extracted using Griffiths technique (6). DNA libraries were prepared by MicrobesNG using Nextera XT Library Prep Kit. Paired-ends (2 × 250 bp) genomic sequencing was carried out by MicrobesNG (Birmingham, UK) using Illumina HiSeq technology. Raw reads were trimmed using Trimmomatic (v0.30) (quality cutoff of Q15) (7), and *de novo* assembly was carried out with SPAdes (v3.7) (8). Genome quality was assessed with CheckM (9). Genome coverage was calculated using BWA, SAMtools (0.1.19), and BEDTools genomecov (2.2.7) (10–12). Genome annotation was performed using the Rapid Annotations Subsystems Technology (RAST) annotation server (v2.0) with the classic RAST pipeline (13). The presence of metabolic pathways was further screened using BlastKoala (2.3) (14).

Bacillus sp. strain X was most closely related to *Bacillus subtilis*, with an average nucleotide identity of 98.26% as determined by ANItools (15). The draft genomes of *Bacillus* sp. strain X yielded 17 contigs, including 2,322 coding sequences (CDSs), 44.92% GC content, and 3,391,284 bp genome size, with 160-fold coverage. *Salarachaeum* sp. strain III X was closely related to Halobacteriales, with 91.36% nucleotide sequence identity. The genome comprised 129 contigs, including 943 CDSs with 67.50% GC content. The genome size was 2,034,915 bp with 30-fold coverage. CheckM analysis identified completion and contamination values of 98.73% and 0.32% for the *Bacillus* genome and 99.28% and 1.52% for the *Salarachaeum* genome.

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

Address correspondence to Michael C. Macey, michael.macey@open.ac.uk, or Ermias Sissay Balcha, ermisislund@gmail.com.

Ermias Sissay Balcha and Michael C. Macey contributed equally to this article. Author order was determined based on seniority.

The authors declare no conflict of interest.

Received 22 December 2023

Accepted 8 February 2024

Published 22 February 2024

Copyright © 2024 Balcha 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/).

Genome mining of *Bacillus* sp. strain X was carried out with antibiotics and Secondary Metabolites Analysis SHell (antiSMASH, v6.0) program (16) and showed the presence of 13 (44.8 %) non-ribosomal peptide gene clusters that encoded a variety of secondary metabolites that produce enzymes, putative antimicrobials, and other bioactive compounds of significant importance in the biotechnological and medical sectors (17). Gene clusters encoding ribosomally synthesized and post-translationally modified peptides and terpenes were also identified from *Salarachaeum* sp. strain III. Regarding biogeochemical cycling, genes for the complete assimilatory sulfate and dissimilatory nitrate pathways were identified in the genome of *Bacillus* sp. strain X, while the genome of *Salarachaeum* sp. strain III contained putative formate dehydrogenase (NAD-dependent) and carbon monoxide dehydrogenase-encoding genes.

ACKNOWLEDGMENTS

We are grateful to the Afar Regional State Mine Resource Development Bureau for granting us research permits and allowing access to the sampling sites. We are also deeply grateful to the Europlanet 2024-RI fund for providing the financial support necessary to conduct this research. Additionally, we would like to thank Research England for their support, particularly through the "Expanding Excellence in England" grant number 124.18.

This research was funded by Research England via the "Expanding Excellence in England" grant number 124.18 and the Europlanet 2024-RI fund.

AUTHOR AFFILIATIONS

¹School of Medical Laboratory Science, College of Health Sciences, Hawassa University, Hawassa, Ethiopia

²Department of Biotechnology, College of Biological and Chemical Engineering, Addis Ababa Science and Technology University, Addis Ababa, Ethiopia

³AstrobiologyOU, School of Environment, Earth and Ecosystem Sciences, The Open University, Milton Keynes, United Kingdom

⁴Dipartimento di Scienze Biologiche, Geologiche e Ambientali, Università di Bologna, Bologna, Italy

⁵Department of Geology, University of Johannesburg, Johannesburg, South Africa

AUTHOR ORCIDs

Ermias Sissay Balcha  <http://orcid.org/0000-0003-3867-0096>

Michael C. Macey  <http://orcid.org/0000-0002-1800-0469>

AUTHOR CONTRIBUTIONS

Ermias Sissay Balcha, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review and editing | Karen Olsson-Francis, Conceptualization, Data curation, Formal analysis, Methodology, Supervision, Validation, Writing – original draft, Writing – review and editing | Ben Stephens, Conceptualization, Formal analysis, Methodology, Writing – original draft, Writing – review and editing | Barbara Cavalazzi, Conceptualization, Methodology, Writing – original draft, Writing – review and editing | Adugna Abdi Woldeesemayat, Conceptualization, Methodology, Writing – original draft, Writing – review and editing | Mesfin Tafesse Gameda, Conceptualization, Writing – original draft, Writing – review and editing | Michael C. Macey, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Supervision, Validation, Writing – original draft, Writing – review and editing

DATA AVAILABILITY

The Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession no. [PRJNA1050408](https://doi.org/10.1093/mra/13.3.0000000) for *Bacillus* sp. strain X and [JAZHPG00000000](https://doi.org/10.1093/mra/13.3.0000000) for *Salarachaeum* sp. strain III. The version described in this paper is the first version.

REFERENCES

1. Abbate E, Bruni P, Sagri M. 2015. Geology of Ethiopia: a review and geomorphological perspectives, p 33–64. In *World geomorphological landscapes*. Springer.
2. Cavalazzi B, Barbieri R, Gómez F, Capaccioni B, Olsson-Francis K, Pondrelli M, Rossi AP, Hickman-Lewis K, Agangi A, Gasparotto G, Glamoclija M, Ori GG, Rodríguez N, Hagos M. 2019. The Dallol geothermal area, Northern Afar (Ethiopia)-an exceptional planetary field analog on earth. *Astrobiology* 19:553–578. <https://doi.org/10.1089/ast.2018.1926>
3. Belilla J, Moreira D, Jardillier L, Reboul G, Benzerara K, López-García JM, Bertolino P, López-Archilla AI, López-García P. 2019. Hyperdiverse archaea near life limits at the polyextreme geothermal Dallol area. *Nat Ecol Evol* 3:1552–1561. <https://doi.org/10.1038/s41559-019-1005-0>
4. Gómez F, Cavalazzi B, Rodríguez N, Amils R, Ori GG, Olsson-Francis K, Escudero C, Martínez JM, Miruts H. 2019. Ultra-small microorganisms in the polyextreme conditions of the Dallol volcano, Northern Afar, Ethiopia. *Sci Rep* 9:7907. <https://doi.org/10.1038/s41598-019-44440-8>
5. Ravimannan N, Pathmanathan S. 2016. Soy flour as alternative culture media for yeasts. Available from: <https://www.researchgate.net/publication/315836118>
6. Griffiths RI, Whiteley AS, O'Donnell AG, Bailey MJ. 2000. Rapid method for coextraction of DNA and RNA from natural environments for analysis of ribosomal DNA- and rRNA-based microbial community composition. *Appl Environ Microbiol* 66:5488–5491. <https://doi.org/10.1128/AEM.66.12.5488-5491.2000>
7. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
8. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>
9. 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>
10. Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841–842. <https://doi.org/10.1093/bioinformatics/btq033>
11. 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>
12. Li H, Durbin R. 2009. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics* 25:1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
13. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, et al. 2008. The RAST server: rapid annotations using subsystems technology. *BMC Genomics* 9:75. <https://doi.org/10.1186/1471-2164-9-75>
14. Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J Mol Biol* 428:726–731. <https://doi.org/10.1016/j.jmb.2015.11.006>
15. Han N, Qiang Y, Zhang W. 2016. ANItools web: a web tool for fast genome comparison within multiple bacterial strains. *Database (Oxford)* 2016:baw084. <https://doi.org/10.1093/database/baw084>
16. Blin K, Shaw S, Kloosterman AM, Charlop-Powers Z, van Wezel GP, Medema MH, Weber T. 2021. AntiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res* 49:W29–W35. <https://doi.org/10.1093/nar/gkab335>
17. Hamdache A, Lamarti A, Aleu J, Collado IG. 2011. Non-peptide metabolites from the genus *Bacillus*. *J Nat Prod* 74:893–899. <https://doi.org/10.1021/np100853e>