

Eight genome sequences of bacterial, environmental isolates from Canada Glacier, Antarctica

Heidi J. Smith,^{1,2} Markus Dieser,^{1,3} Christine M. Foreman^{1,3}

AUTHOR AFFILIATIONS See affiliation list on p. 2.

ABSTRACT Sediments in cryoconite holes and meltwater streams in the McMurdo Dry Valleys, Antarctica, provide both substrates and conditions that support life in an arid polar desert. Here, we report the genomic sequences of eight environmental, bacterial isolates from Canada Glacier cryoconite holes and stream. These isolates span three major phyla.

KEYWORDS Antarctica, glacier, cryoconite hole, stream

In the barren polar desert of the McMurdo Dry Valleys, Antarctica, meltwater-driven supraglacial depressions (i.e., cryoconite holes) and streams such as those in the Canada Glacier basin are among the top refuges for microorganisms (1–6). Cryoconite holes are cylindrical, water-filled depressions on glacial surfaces formed by the deposition and accumulation of aeolian material. The Canada Stream is an ephemeral stream, fed by runoff from the Canada Glacier.

While microbes thrive in cryoconite holes and the stream channel (7–10), rapid response and adaptation are required to withstand frequent disturbances, freezing, desiccation, and transition to quiescence. Our goal is to understand the genetic traits linked to persistence under the multitude of environmental stressors in these systems.

Cryoconite and stream water samples from the Canada Glacier (77°37'S, 162°59'E) were collected in December 2009 and are described elsewhere (9). Bacteria were isolated aerobically on R2A agar plates at 4°C. The cetyltrimethylammonium bromide procedure was used for extracting DNA from bacterial isolates grown to late exponential phase in R2A broth at 4°C while shaking at 150 rpm (11). Standard quality shotgun libraries for bacterial strains CAN_C2, CAN_C3, CAN_C7, CAN_S1, CAN_S2, CAN_S4, and CAN_S7 were sequenced on the Illumina NovaSeq S4 platform (2 × 151 bp paired-end reads) (12). Libraries were prepared on the PerkinElmer Sciclone NGS robotic liquid handling workstation using the Kapa Biosystems Library Kit. DNA was sheared to 459 bp using a Covaris LE220-focused ultrasonicator. DNA fragments were size selected by double solid phase reversible immobilization (SPRI). Fragments were end repaired, A tailed, and ligated with Illumina compatible sequencing adaptors. Raw Illumina sequences were quality filtered using BBTools v38.95 (QV ≥20; length ≥100 bases) (13) per JGI SOP 1061 and assembled with SPAdes (≥version v3.14.1; –phred-offset 33 –cov-cutoff auto –t 16 m 64 –careful –k 25,55,95) (14). Contigs with a length <1 kb (BBTools reformat.sh: minlength = 1,000 ow = t) were discarded.

For improved high-quality draft genomes, PacBio SMRTbell libraries for bacterial strain CAN_C5 were sequenced on Pacific Biosciences (PacBio) Sequel platform (15). PacBio sequencing generated 889,350 reads with an average length of 8,908 bp. DNA was sheared around 10 kb using Megaruptor 3 (Diagenode). Sheared DNA was treated with exonuclease, DNA repair enzyme mix, end-repair/A-tailing mix, and ligated with barcoded overhang adapters using SMRTbell Express Template Prep Kit 2.0 (PacBio). Libraries were purified with AMPure PB Beads (PacBio) and bound to

Editor Julia A. Maresca, SUNY College of Environmental Science and Forestry, Syracuse, New York, USA

Address correspondence to Christine M. Foreman, cforeman@montana.edu.

The authors declare no conflict of interest.

See the funding table on p. 3.

Received 27 November 2023

Accepted 19 June 2024

Published 11 July 2024

Copyright © 2024 Smith 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/).

TABLE 1 Summary of genome characteristics^a

Platform	Organism (% match to genus) Accession number	Coverage (x)	# Contigs	L50/N50 (bp)	Size (bp)	GC (%)	# Genes	# C
Illumina	<i>Cryobacterium</i> sp. CAN_C2 (99.5) JADOUX000000000 SRR13292036	419.2	58	8/149,434	3,600,773	64.66	3,541	– ^b
Illumina	<i>Cryobacterium</i> sp. CAN_C3 (99.4) JAXCHQ000000000 SRR25764242	394.8	60	9/149,434	3,604,084	64.65	3,519	–
PACBIO	<i>Arthrobacter</i> sp. CAN_C5 (99.6) JAGGMZ000000000 SRR26356106	250.1	1	1/3,860,921	3,860,921	64.15	3,796	–
Illumina	<i>Mycetocola</i> sp. CAN_C7 (97.8) JADOUY000000000 SRR13164828	374.4	20	3/556,454	3,420,744	65.49	3,308	–
Illumina	<i>Janthinobacterium</i> sp. CAN_S1 (99.7) JADOUZ000000000 SRR13164876	252.0	72	10/151,624	4,870,004	60.87	4,452	–
Illumina	<i>Flavobacterium</i> sp. CAN_S2 (99.6) JADOVA000000000 SRR13164877	197.8	17	2/469,638	3,844,435	34.22	3,572	2
Illumina	<i>Salinibacterium</i> sp. CAN_S4 (99.6) JADOVB000000000 SRR13164888	384.9	11	2/547,050	3,068,493	65.19	3,035	1
Illumina	<i>Janthinobacterium</i> sp. CAN_S7 (98.4) JBANDI000000000 SRR25764251	328.9	68	12/134,133	5,015,770	60.65	4,645	–

^aC = CRISPR.

^b–, none.

Sequel II polymerase 2.0 using the Sequel II Binding Kit 2.0. Libraries were sequenced using tbd-sample-dependent sequencing primers, 8M v1 SMRT cells, and Version 2.0 sequencing chemistry with 1 × 900 sequencing movie run times. Reads >5 kb were assembled with hierarchical genome assembly process (HGAP) [smrtlink/8.0.0.80529, HGAP 4 (1.0)] using default settings (16).

CRISPR elements were identified using the program CRT (CRISPR Recognition Tool) (17) or PILER-CR (18). All genomes were annotated in the Integrated Microbial Genomes (IMG) database (≥IMG Annotation Pipeline v.5.0.19) (19). Genome completeness was estimated using checkM (20). The genome-sequence-acquired 16S ribosomal RNA gene was queried in command line against the SILVA database 138.1 with blastn, from BLAST+ 2.13.0 for taxonomic classification (21). Sequence details are given in Table 1.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation under grant ANT-0838970 to C.M.F., NASA Fellowship NESSF Planet11R-0011 to H.J.S., and ANT-2037963 to H.J.S., M.D., and C.M.F. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported under Contract No. DE-AC02-05CH11231. Data were generated for JGI Proposal #505037.

AUTHOR AFFILIATIONS

¹Center for Biofilm Engineering, Montana State University, Bozeman, Montana, USA

²Department of Microbiology and Cell Biology, Montana State University, Bozeman, Montana, USA

³Department of Chemical & Biological Engineering, Montana State University, Bozeman, Montana, USA

AUTHOR ORCIDS

Heidi J. Smith  <http://orcid.org/0000-0001-7234-9146>

Markus Dieser  <http://orcid.org/0000-0001-8539-6646>

Christine M. Foreman  <http://orcid.org/0000-0003-0230-4692>

FUNDING

Funder	Grant(s)	Author(s)
National Science Foundation (NSF)	ANT-0838970	Christine M. Foreman
National Science Foundation (NSF)	ANT-2037963	Christine M. Foreman Heidi J. Smith Markus Dieser
National Aeronautics and Space Administration (NASA)	NESSF Planet11R-0011	Heidi J. Smith

AUTHOR CONTRIBUTIONS

Heidi J. Smith, Conceptualization, Writing – original draft | Markus Dieser, Conceptualization, Writing – review and editing | Christine M. Foreman, Conceptualization, Funding acquisition, Writing – review and editing

DATA AVAILABILITY

The JGI genome project (ID: 505037) is available from the IMG database. The genome sequences have been submitted to GenBank under the accession numbers listed in Table 1.

REFERENCES

- Wlostowski AN, Gooseff MN, McKnight DM, Jaros C, Lyons WB. 2016. Patterns of hydrologic connectivity in the McMurdo Dry Valleys, Antarctica: a synthesis of 20 years of hydrologic data. *Hydrol Processes* 30:2958–2975. <https://doi.org/10.1002/hyp.10818>
- Fountain AG, Tranter M, Nylen TH, Lewis KJ, Mueller DR. 2004. Evolution of cryoconite holes and their contribution to meltwater runoff from glaciers in the McMurdo Dry Valleys, Antarctica. *J Glaciol* 50:35–45. <https://doi.org/10.3189/172756504781830312>
- Gooseff MN, McKnight DM, Doran PT, Fountain A. 2022. Long - term stream hydrology and meteorology of a Polar Desert, the McMurdo Dry Valleys, Antarctica. *Hydrol Process* 36:e14623. <https://doi.org/10.1002/hyp.14623>
- McKnight DM, Tate CM, Andrews ED, Niyogi DK, Cozzetto K, Welch K, Lyons WB, Capone DG. 2007. Reactivation of a cryptobiotic stream ecosystem in the McMurdo Dry Valleys, Antarctica: a long-term geomorphological experiment. *Geomorphology* 89:186–204. <https://doi.org/10.1016/j.geomorph.2006.07.025>
- Kohler TJ, Stanish LF, Crisp SW, Koch JC, Liptzin D, Baeseman JL, McKnight DM. 2015. Life in the main channel: long-term hydrologic control of microbial mat abundance in McMurdo Dry Valley streams, Antarctica. *Ecosystems* 18:310–327. <https://doi.org/10.1007/s10021-014-9829-6>
- Fountain AG, Nylen TH, Tranter M, Bagshaw E. 2008. Temporal variations in physical and chemical features of cryoconite holes on Canada glacier, McMurdo Dry Valleys, Antarctica. *J Geophys Res* 113:G1. <https://doi.org/10.1029/2007JG000430>
- McKnight DM, Niyogi DK, Alger AS, Bomblies A, Conovitz PA, Tate CM. 1999. Dry valley streams in Antarctica: ecosystems waiting for water. *BioScience* 49:985. <https://doi.org/10.2307/1313732>
- Van Horn DJ, Wolf CR, Colman DR, Jiang X, Kohler TJ, McKnight DM, Stanish LF, Yazzie T, Takacs-Vesbach CD. 2016. Patterns of bacterial biodiversity in the glacial meltwater streams of the McMurdo Dry Valleys, Antarctica. *FEMS Microbiol Ecol* 92:fiw148. <https://doi.org/10.1093/femsec/fiw148>
- Smith HJ, Dieser M, McKnight DM, SanClements MD, Foreman CM. 2018. Relationship between dissolved organic matter quality and microbial community composition across polar glacial environments. *FEMS Microbiol Ecol* 94:fiy090. <https://doi.org/10.1093/femsec/fiy090>
- Foreman CM, Sattler B, Mikucki JA, Porazinska DL, Priscu JC. 2007. Metabolic activity and diversity of cryoconites in the Taylor Valley, Antarctica. *J Geophys Res* 112:G4. <https://doi.org/10.1029/2006JG000358>
- William S, Feil H, Copeland A. 2012. Bacterial genomic DNA isolation using CTAB. *Sigma* 50:6876.
- Bennett S. 2004. Solexa Ltd. *Pharmacogenomics* 5:433–438. <https://doi.org/10.1517/14622416.5.4.433>
- Bushnell B. BBTools software package (version 38.95). <https://bbtools.jgi.doe.gov>.
- 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>
15. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, et al. 2009. Real-time DNA sequencing from single polymerase molecules. *Science* 323:133–138. <https://doi.org/10.1126/science.1162986>
 16. Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* 10:563–569. <https://doi.org/10.1038/nmeth.2474>
 17. Bland C, Ramsey TL, Sabree F, Lowe M, Brown K, Kyrpides NC, Hugenholtz P. 2007. CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics* 8:1–8. <https://doi.org/10.1186/1471-2105-8-209>
 18. Edgar RC. 2007. PILER-CR: fast and accurate identification of CRISPR repeats. *BMC Bioinformatics* 8:1–6. <https://doi.org/10.1186/1471-2105-8-18>
 19. Chen I-M, Chu K, Palaniappan K, Pillay M, Ratner A, Huang J, Huntemann M, Varghese N, White JR, Seshadri R, Smirnova T, Kirton E, Jungbluth SP, Woyke T, Elie-Fadrosh EA, Ivanova NN, Kyrpides NC. 2019. IMG/M v. 5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes. *Nucleic Acids Res* 47:D666–D677. <https://doi.org/10.1093/nar/gky901>
 20. 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>
 21. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 41:D590–D596. <https://doi.org/10.1093/nar/gks1219>