Made available through Montana State University's ScholarWorks

Fourteen genome sequences from bacterial, environmental isolates from the Cotton Glacier stream, Antarctica

Christine M. Foreman, Heidi J. Smith, Markus Dieser

Accessibility Disclaimer:

For a more accessible version of this document, please submit an accessibility request form through the Montana State University Library website.

AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

Check for updates

| Environmental Microbiology | Announcement

Fourteen genome sequences from bacterial, environmental isolates from the Cotton Glacier stream, Antarctica

Christine M. Foreman,[1,2] **Heidi J. Smith,**[1,3] **Markus Dieser**[1,2]

AUTHOR AFFILIATIONS See affiliation list on p. 3.

ABSTRACT We report the genomic sequences of 14 bacterial isolates from a supraglacial stream on the Cotton Glacier, Antarctica. Fine sediments in the streambed provide habitat for bacterial growth and biofilm formation. The stream represents a natural laboratory for studying the evolution and adaptation of microbes to a humic-free environment.

KEYWORDS supraglacial, Antarctica, biofilm

The Cotton Glacier stream (77°07′S, 161°40′E), Antarctica, is a ~16 km long sediment-laden, supraglacial stream. Detailed studies have highlighted its biogeochemical processes and downstream ramifications (1–5). This stream represents a natural laboratory for studying the evolution and adaptation of microbes to the absence of humics (1). Stream water was collected during the austral summers of 2009 and is described elsewhere (1). Bacteria were isolated aerobically on R2A agar plates at 4°C in the dark. Isolates were selected and cultured based on colony morphology. Isolates stored in 25% glycerol at −80°C were grown on R2A agar plates at 4°C for 7 days. Single colonies were then inoculated in R2A broth (20 mL) at 4°C while shaking at 150 rpm to the late exponential phase. The cetyltrimethylammonium bromide procedure was used for extracting DNA (6).

Standard quality shotgun libraries for bacterial strains CG_S6, CG_S7, CG_23.3, CG_9.1, CG_9.2, CG_9.5, CG_9.7, CG_9.8, CG_9.10, and CG_9.11 were sequenced on the Illumina NovaSeq S4 platform (2 × 151 bp paired-end reads) (7). 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-SPRI. Fragments were end-repaired, A-tailed, and ligated with Illumina-compatible sequencing adaptors. Raw Illumina sequences were quality filtered using BBTools (8) 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) (9). Contigs with a length <1 kb (BBTools ≥ v.38.86, reformat.sh: minlength = 1,000 ow = t) were discarded.

For improved high-quality draft genomes, PacBio SMRTbell libraries for bacterial strains CG_23.4, CG_23.5, CG_23.6, and CG_9.6 were sequenced on the Pacific Biosciences (PacBio) Sequel II platform (10). 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 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. Additionally, collapsed adapters were removed with BBTools v38.87. Reads >5 kb from CG_23.5 and CG_9.6 were assembled with the hierarchical genome assembly process

Editor Simon Roux, DOE Joint Genome Institute, Berkeley, California, 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 7 February 2025
Accepted 28 July 2025
Published 13 August 2025

Copyright © 2025 Foreman et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Downloaded from https://journals.asm.org/journal/mra on 17 April 2026 by 153.90.170.56.

TABLE 1 Summary of genome characteristics[a,b]

Platform	Organism [% match to genus, accession #]] Accession number	# reads Coverage (×) Completeness (%) Contamination (%)	# Contigs	L_{50}/N_{50} (bp)	Size (bp)	GC (%)	# Genes	# C	# P
Illumina	*Janthinobacterium* sp. CG_S6 [98.1; NR_132608.1] JAXCHS000000000 SRR25764247	13,218,236 246.4 99.26 0.19	79	12/181,542	6,110,077	65.69	5,505	2	–
Illumina	*Pedobacter* sp. CG_S7 [98.2; NR_148294.1] JBEONS000000000 SRR28961170	1,691,946 405.4 99.84 0.96	108	12/93,633	3,854,172	37.16	3,582	2	–
Illumina	*Janthinobacterium* sp. CG_23.3 [98.1; NR_132608.1] JBJXIB000000000 SRR30853869	8,213,898 256.4 98.46 0.35	280	51/38,360	6,116,288	65.55	5,722	2	–
PACBIO	*Janthinobacterium* sp. CG_23.4 [99,4; NR_132608.1] JARXUX000000000 SRR12927869	905,709 193.2 99.95 0.0	2	1/4,988,477	5,012,930	60.74	4,596	–	1
PACBIO	*Flavobacterium* sp. CG_23.5 [98.4; NR_043891.1] JAGGNA000000000 SRR12927877	396,643 261.4 98.94 0.0	1	1/3,672,252	3,672,252	33.93	3,365	1	–
PACBIO	*Polaromonas* sp. CG_23.6 [99.2; NR_109013.1] JARXUY000000000 SRR12927893	637,596 214.6 99.66 0.25	4	1/4,099,184	4,511,363	60.01	4,444	–	3
Illumina	*Flavobacterium* sp. CG_9.1 [99.8; NR_113714.1] JADOVC000000000 SRR13164898	5,107,940 201.4 98.94 2.68	86	12/111,844	3,938,215	34.35	3,644	1	–
Illumina	*Polaromonas* sp. CG_9.2 [99.2; NR_109013.1] JADOVF000000000 SRR13164963	8,771,292 323.6 99.66 0.25	55	8/192,115	4,152,851	60.1	4,012	–	–
Illumina	*Polaromonas* sp. CG_9.5 [99.3; NR_109013.1] JAXCHR000000000 SRR25764248	6,678,012 390.8 99.73 0.0	76	9/157,676	3,803,867	60.18	3,659	–	–
PACBIO	*Cryobacterium* sp. CG_9.6 [99.3; NR_042170.1] JARXUZ000000000 SRR12927894	659,510 271.1 99.24 0.51	2	1/3,352,313	3,539,656	63.32	3,434	2	1
Illumina	*Polaromonas* sp. CG_9.7 [99.2; NR_109013.1] JADOVG000000000 SRR13164962	1,787,966 66.5 99.66 0.25	55	8/192,115	4,152,641	60.1	4,007	–	–
Illumina	*Frigoribacterium* sp. CG_9.8 [97.9; NR_026511.1 JADOVH000000000 SRR13164979	9,006,494 475.3 99.49 0.51	11	2/614,762	2,845,437	63.3	2,825	1	–

(*Continued on next page*)

Downloaded from https://journals.asm.org/journal/mra on 17 April 2026 by 153.90.170.56.

TABLE 1 Summary of genome characteristics[a,b] (*Continued*)

Platform	Organism [% match to genus, accession #]] Accession number	# reads Coverage (×) Completeness (%) Contamination (%)	# Contigs	L_{50}/N_{50} (bp)	Size (bp)	GC (%)	# Genes	# C	# P
Illumina	*Flavobacterium* sp. CG_9.10 [99.3; NR_199451.1] JADOVD000000000 SRR13164902	4,286,438 191.7 98.94 0.0	43	7/210,281	3,529,796	33.79	3,284	1	–
Illumina	*Polaromonas* sp. CG_9.11 [99.2; NR_109013.1] JADOVE000000000 SRR13164964	7,279,460 270.2 99.61 0.03	77	10/145,329	4,267,931	61.41	4,191	–	–

[a]C = CRISPR, P = Plasmid.
[b]–, none.

[HGAP] (smrtlink/8.0.0.80529, HGAP 4 (1.0)) using default settings (11). Sequence reads from isolates CG_23.4 and CG_9.6 were assembled with Flye 2.8.1 (–asm-coverage 100 –plasmid –genome-size 5000000 t 64 –pacbio-raw (fasta list)) (12). Contigs with a probability of being a plasmid were identified using TensorFlow (13).

CRISPR elements were identified using the programs CRT v.1.8.2 (CRISPR Recognition Tool) (14) or PILER-CR (15). All genomes were annotated in the Integrated Microbial Genomes (IMG) database (≥IMG Annotation Pipeline v.5.1.2) (16). Genome completeness was estimated using checkM v.1.2.2 (17). 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 (18). Sequence details are given in Table 1.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation under grants ANT-0838970 to C.M.F., NASA Fellowship NESSF Planet 11R-0011 to H.J.S. and NSF 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. These data were generated for JGI Proposal #505037.

AUTHOR AFFILIATIONS

[1]Center for Biofilm Engineering, Montana State University, Bozeman, Montana, USA
[2]Department of Chemical & Biological Engineering, Montana State University, Bozeman, Montana, USA
[3]Department of Microbiology and Cell Biology, Montana State University, Bozeman, Montana, USA

AUTHOR ORCIDs

Christine M. Foreman http://orcid.org/0000-0003-0230-4692
Heidi J. Smith http://orcid.org/0000-0001-7234-9146

FUNDING

Funder	Grant(s)	Author(s)
National Science Foundation	ANT-0838970	Christine M. Foreman

Downloaded from https://journals.asm.org/journal/mra on 17 April 2026 by 153.90.170.56.

Funder	Grant(s)	Author(s)
National Aeronautics and Space Administration	NESSF Planet 11R-0011	Heidi J. Smith
National Science Foundation	NSF 2037963	Christine M. Foreman Heidi J. Smith

AUTHOR CONTRIBUTIONS

Christine M. Foreman, Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – review and editing | Heidi J. Smith, Conceptualization, Investigation, Methodology, Writing – review and editing | Markus Dieser, Conceptualization, Investigation, Methodology, writing - original draft, Writing – review and editing

DATA AVAILABILITY

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. These data were generated for JGI Proposal #505037. The genome sequences have been submitted to GenBank and are available through the accession numbers listed in Table 1.

REFERENCES

1. Foreman CM, Cory RM, Morris CE, SanClements MD, Smith HJ, Lisle JT, Miller PL, Chin YP, McKnight DM. 2013. Microbial growth under humic-free conditions in a supraglacial stream system on the cotton glacier, Antarctica. Environ Res Lett 8:035022. https://doi.org/10.1088/1748-9326/8/3/035022
2. 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
3. Smith HJ, Tigges M, D'Andrilli J, Parker A, Bothner B, Foreman CM. 2018. Dynamic processing of DOM: insight from exometabolomics, fluorescence spectroscopy, and mass spectrometry. Limnol Oceanogr Lett 3:225–235. https://doi.org/10.1002/lol2.10082
4. Smith HJ, Foster RA, McKnight DM, Lisle JT, Littmann S, Kuypers MMM, Foreman CM. 2017. Microbial formation of labile organic carbon in Antarctic glacial environments. Nature Geosci 10:356–359. https://doi.org/10.1038/ngeo2925
5. SanClements MD, Smith HJ, Foreman CM, Tedesco M, Chin YP, Jaros C, McKnight DM. 2017. Biogeophysical properties of an expansive Antarctic supraglacial stream. Antarct Sci 29:33–44. https://doi.org/10.1017/S0954102016000456
6. William S, Feil H, Copeland A. 2012. Bacterial genomic DNA isolation using CTAB. Sigma 50:6876.
7. Bennett S. 2004. Solexa Ltd. Pharmacogenomics 5:433–438. https://doi.org/10.1517/14622416.5.4.433
8. Bushnell B. 2017 BBTools software package. Available from: https://bbtools.jgi.doe.gov
9. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski 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
10. 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
11. 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
12. 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
13. Rampasek L, Goldenberg A. 2016. TensorFlow: biology's gateway to deep learning? Cell Syst 2:12–14. https://doi.org/10.1016/j.cels.2016.01.009
14. 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
15. 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
16. Chen I-MA, 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, Eloe-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
17. 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
18. 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:D590–6. https://doi.org/10.1093/nar/gks1219

Downloaded from https://journals.asm.org/journal/mra on 17 April 2026 by 153.90.170.56.