Title
Metagenome-Assembled Genomes from Appalachian Acid Mine Drainage Sites.

Permalink
https://escholarship.org/uc/item/2k41r4v8

Journal
Microbiology resource announcements, 12(5)

ISSN
2576-098X

Authors
Williams, Caden
Macalady, Jennifer
Grettenberger, Christen

Publication Date
2023-05-01

DOI
10.1128/mra.01280-22

Peer reviewed

Metagenome-Assembled Genomes from Appalachian Acid Mine Drainage Sites

Caden Williams,[a] Jennifer Macalady,[b] Christen Grettenberger[a,c]

aDepartment of Earth and Planetary Sciences, University of California, Davis, California, USA
bDepartment of Geosciences, Pennsylvania State University, University Park, Pennsylvania, USA
cDepartment of Ecotoxicology, University of California, Davis, California, USA

ABSTRACT   Here, we report 7 metagenome-assembled genomes (MAGs) isolated from acid mine drainage sites in the eastern United States. Three genomes are *Archaea*, including two from the phylum *Thermoproteota* and one from *Euryarchaeota*. Four genomes are bacterial, with one from the phylum *Candidatus* Eremiobacteraeota (formerly WPS-2), one from *Acidimicrobiales* (*Actinobacteria*), and two from *Gallionellaceae* (*Proteobacteria*).

Acid mine drainage (AMD) sites are unique environments due to their low pH and high metal concentrations. Although many organisms are intolerant of these conditions, bacterial and archaeal acidophiles inhabit AMD sites (1–5). AMD is a major environmental issue and creates extreme ecosystems where microbial communities drive biogeochemical cycling (3, 6–10). For example, some species oxidize metal sulfides or dissolved metal species, facilitating acidification (9, 11–14). However, few genomes are available from AMD environments, especially from newly discovered lineages. This genome report provides genome sequences from 7 AMD populations which provide insight into the microbial activity and diversity in AMD.

Samples were collected from four AMD sites across the Appalachian region, including Porcupine Run in Pennsylvania (41.34378 N, 78.37827 W), Martin Run in West Virginia (39.550828 N, 79.639311 W), and Maxine Mine Geological Overburden Pile (GOB) (33.57420 N, 87.15634 W) and Simpson (33.56417 N, 87.17316 W) in Alabama. Approximately 3 grams of surface sediment was collected in a sterile tube, preserved with 15 mL of RNALater, transported on ice, and stored at −20°C. DNA was extracted using a PowerBiofilm DNA extraction kit (MoBio, Carlsbad, USA). Each sample was extracted twice with a 2- and 4-min vortex adaptor bead beating time, and the products were pooled. Libraries were prepared using the Nextera XT library preparation kit following the manufacturer's instructions. Paired-end 2×150-bp sequencing was performed using an Illumina HiSeq 2500 instrument. Dissolved oxygen, pH, and oxidation reduction potential (ORP) were measured using a Hach HQ40d portable meter. Ferrous iron ([Fe2+]) was measured using a Hach DR 2700 spectrophotometer (Hach, Loveland, USA) (Table 1).

Data processing and analysis were performed in KBase using default parameters (15). Trimmomatic (v.0.36) removed adaptors and low-quality regions (16). The quality of the trimmed data was checked using FastQC (v.0.11.9) (17). Reads were assembled using MEGAHIT (v.1.2.9) (18). Contigs were binned with MetaBAT2 contig binning (v.1.7) (19). Bins were filtered by quality using CheckM (v.1.0.18) for completeness of ≥90% and contamination of ≤5% (20). Bins were classified with GTDB-Tk (v.1.7.0) which assigns genomes to domains using marker genes and places genomes into a reference tree using domain-specific markers (21). Genomes were annotated with Prokka (v.1.14.5) (22). Genome coverage was calculated by mapping the trimmed reads to the MAG using Bowtie 2 (v.2.3.2) (23).

We recovered 7 MAGs. The MAGs were 94.7 to 99.5% complete with 0.8 to 3.2% contamination and contained 1.7 to 2.8 Mbp each (Table 1). GC content was 37.9 to 63.1%.

TABLE 1 Summary of site geochemistry, metagenome statistics, and MAG properties

Site	Site geochemistry				Metagenome statistics				Genome statistics									
	pH	ORP	[Fe] (mg/L)	Dissolved oxygen (mg/L)	Accession no.	No. of quality-controlled reads	No. of contigs	N_{50} (bp)	Accession no.	Genome name	Classification	Total length (Mb)	GC (%) content	No. of contigs	Completeness (%)	Contamination (%)	Coverage (x)	No. of protein-coding genes
Porcupine Run	4.2	261.1	36.5	7.2	SAMN32007064	3,671,780	2,946	7,242	SAMN32097021	Porcupine6	Bacteria; Candidatus Eremiobacteraeota; Eremiobacteria; Baltobacterales; Ca. Baltobacteraceae; Ca. Rubrimentiphilum; JABEUX01 sp01304 4185	2.41	62.3	184	94.5	2.01	16.2	2,433
									SAMN32097022	Porcupine7	Bacteria; Actinobacteriota; Acidimicrobiia; Acidimicrobiales; Acidimicrobiaceae; RAAP-2	2.48	63.1	146	98.29	2.99	24.0	2,422
Martin Run	3.8	477	7.8	9.3	SAMN32007066	7,611,764	700	7,575	SAMN32097023	Martin2	Archaea; Thermoproteota; Nitrososphaeria; Nitrososphaerales; Ca. Nitrosotalea	1.7	37.9	97	99.51	0.97	11.7	2,016
Maxine Mine GOB	4.0	134.5	199	0.4	SAMN32007065	6,021,450	2,056	5,836	SAMN32097024	Maxine3	Archaea; Thermoproteota; Nitrososphaeria; Nitrososphaerales; CADDZS01; CADDZS01	2.78	44.3	237	94.66	3.16	17.0	3,119
									SAMN32097025	Maxine5	Archaea; Euryarchaeota; Thermoplasmata; Methanomassiliicoccales	1.76	51.2	131	95.81	0.8	12.1	1,767
Simpson	7.3	−84.7	1.3	0.7	SAMN32007067	1,675,384	1,071	9,715	SAMN32097019	Simpson2	Bacteria; Proteobacteria; Gammaproteobacteria; Burkholderiales; Gallionellaceae; Gallionella	1.97	51.4	101	97.14	0.99	11.3	1,985
									SAMN32097020	Simpson3	Bacteria; Proteobacteria; Gammaproteobacteria; Burkholderiales; Gallionellaceae; Sideroxydans	2.21	55.2	241	94.5	2.76	8.7x	2,260

Genomes were contained in 97 to 241 contigs. Three genomes were archaeal, including two from the order *Nitrososphaerales* and one from *Thermoplasmatales*. Four genomes were bacterial, including one from the family *Candidatus* Baltobacteraceae within the recently described phylum *Candidatus* Eremiobacteraeota (formerly WPS-2), one from the actinobacterial order *Acidimicrobiales*, and two from the family *Gallionellaceae*.

Data availability. Sequence files are available under BioProject PRJNA907387. Unassembled metagenomic sequences are available under accessions SRX18535480 to SRX18535477, and draft genomes are available under accessions SAMN32097019 to SAMN32097025.

REFERENCES

1. Druschel GK, Baker BJ, Gihring TM, Banfield JF. 2004. Acid mine drainage biogeochemistry at Iron Mountain, California. Geochem Trans 5:13. https://doi.org/10.1186/1467-4866-5-13.
2. Johnson DB, Hallberg KB. 2005. Acid mine drainage remediation options: a review. Sci Total Environ 338:3–14. https://doi.org/10.1016/j.scitotenv.2004.09.002.
3. Johnson DB. 1995. Acidophilic microbial communities: candidates for bioremediation of acidic mine effluents. Int Biodeter Biodegr 35:41–58. https://doi.org/10.1016/0964-8305(95)00065-D.
4. Hua Z-S, Han Y-J, Chen L-X, Liu J, Hu M, Li S-J, Kuang J-L, Chain PSG, Huang L-N, Shu W-S. 2015. Ecological roles of dominant and rare prokaryotes in acid mine drainage revealed by metagenomics and metatranscriptomics. ISME J 9:1280–1294. https://doi.org/10.1038/ismej.2014.212.
5. Bao Y, Guo C, Wang H, Lu G, Yang C, Chen M, Dang Z. 2016. Fe- and S-metabolizing microbial communities dominate an AMD-contaminated river ecosystem and play important roles in Fe and S cycling. Geomicrobiol J 34:695–705. https://doi.org/10.1080/01490451.2016.1243596.
6. Kuang J-L, Huang L-N, Chen L-X, Hua Z-S, Li S-J, Hu M, Li J-T, Shu W-S. 2013. Contemporary environmental variation determines microbial diversity patterns in acid mine drainage. ISME J 7:1038–1050. https://doi.org/10.1038/ismej.2012.139.
7. Jones DS, Kohl C, Grettenberger C, Larson LN, Burgos WD, Macalady JL. 2015. Geochemical niches of iron-oxidizing acidophiles in acidic coal mine drainage. Appl Environ Microbiol 81:1242–1250. https://doi.org/10.1128/AEM.02919-14.
8. Hallberg KB, Coupland K, Kimura S, Johnson DB. 2006. Macroscopic streamer growths in acidic, metal-rich mine waters in North Wales consist of novel and remarkably simple bacterial communities. Appl Environ Microbiol 72:2022–2030. https://doi.org/10.1128/AEM.72.3.2022-2030.2006.
9. Grettenberger CL, Hamilton TL. 2021. Metagenome-assembled genomes of novel taxa from an acid mine drainage environment. Appl Environ Microbiol 87:e0077221. https://doi.org/10.1128/AEM.00772-21.
10. Baker BJ, Banfield JF. 2003. Microbial communities in acid mine drainage. FEMS Microbiol Ecol 44:139–152. https://doi.org/10.1016/S0168-6496(03)00028-X.
11. Jones RM, Johnson DB. 2015. Acidithrix ferrooxidans gen. nov., sp. nov.; a filamentous and obligately heterotrophic, acidophilic member of the Actinobacteria that catalyzes dissimilatory oxido-reduction of iron. Res Microbiol 166:111–120. https://doi.org/10.1016/j.resmic.2015.01.003.
12. Sheng Y, Kaley B, Bibby K, Grettenberger C, Macalady JL, Wang G, Burgos WD. 2017. Bioreactors for low-pH iron (ii) oxidation remove considerable amounts of total iron. RSC Adv 57:35962–35972. https://doi.org/10.1039/C7RA03717A.
13. Dall'Agnol H, Ñancucheo I, Johnson DB, Oliveira R, Leite L, Pylro VS, Holanda R,

Grail B, Carvalho N, Nunes GL, Tzotzos G, Fernandes GR, Dutra J, Orellana SC, Oliveira G. 2016. Draft genome sequence of "Acidibacillus ferrooxidans" ITV01, a novel acidophilic firmicute isolated from a chalcopyrite mine drainage site in Brazil. Genome Announc 4:e01748-15. https://doi.org/10.1128/genomeA.01748-15.
14. Johnson DB, Hallberg KB, Hedrich S. 2014. Uncovering a microbial enigma: isolation and characterization of the streamer-generating, iron-oxidizing, acidophilic bacterium "Ferrovum myxofaciens". Appl Environ Microbiol 80:672–680. https://doi.org/10.1128/AEM.03230-13.
15. Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, Dehal P, Ware D, Perez F, Canon S, Sneddon MW, Henderson ML, Riehl WJ, Murphy-Olson D, Chan SY, Kamimura RT, Kumari S, Drake MM, Brettin TS, Glass EM, Chivian D, Gunter D, Weston DJ, Allen BH, Baumohl J, Best AA, Bowen B, Brenner SE, Bun CC, Chandonia J-M, Chia J-M, Colasanti R, Conrad N, Davis JJ, Davison BH, DeJongh M, Devoid S, Dietrich E, Dubchak I, Edirisinghe JN, Fang G, Faria JP, Frybarger PM, Gerlach W, Gerstein M, Greiner A, Gurtowski J, Haun HL, He F, Jain R, et al. 2018. KBase: the United States Department of Energy Systems Biology Knowledgebase. Nat Biotechnol 36:566–569. https://doi.org/10.1038/nbt.4163.
16. 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.
17. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
18. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. 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.
19. Kang DD, Froula J, Egan R, Wang Z. 2015. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ 3:e1165. https://doi.org/10.7717/peerj.1165.
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. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2019. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics 36:1925–1927. https://doi.org/10.1093/bioinformatics/btz848.
22. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153.
23. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923.