










Metagenome-Assembled Genome Sequences of Novel Prokaryotic Species from the Mercury-Contaminated East Fork Poplar Creek, Oak Ridge, Tennessee, USA

 Minjae Kim,^a  Regina L. Wilpiszkeski,^a  Michael Wells,^a  Ann M. Wymore,^a  Caitlin M. Gionfriddo,^a  Scott C. Brooks,^b
 Mircea Podar,^a  Dwayne A. Elias^a

^aBiosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA

^bEnvironmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA

ABSTRACT We sequenced two metagenomes of sediments from the East Fork Poplar Creek in the Oak Ridge Reservation (Oak Ridge, TN), a natural stream that has been contaminated with Hg from upstream sources, and we reconstructed 28 metagenome-assembled genomes of novel prokaryotic species.

The East Fork Poplar Creek (EFPC) ecosystem is contaminated with Hg from its use to separate Li isotopes for thermonuclear weapons development during the 1950s and early 1960s at the Y-12 National Security Complex in Oak Ridge, Tennessee (1). The release of Hg into the EFPC ecosystem has selected for a unique microbial community that includes many Hg methylators (2). In a recent study, we described the microbial community using both 16S rRNA and shotgun metagenomic sequencing of eight sediment cores obtained along the EFPC (3). Here, we have expanded and enhanced this work by identifying 28 metagenome-assembled genomes (MAGs) from a new sediment core.

The core was collected downstream in the New Horizon Region of the EFPC in January 2018 and generated two samples; one (EFPC-H) was from a depth of 0 to 3 cm, and the other (EFPC-K) was from a depth of 9 to 12 cm. Each core section was homogenized prior to the DNA extraction. The DNA was extracted from ~0.25 g of sediment using the MoBio PowerSoil DNA kit, and the DNA was quantified using a Qubit 2.0 fluorometer (Life Technologies, Carlsbad, CA, USA). DNA library preparation and sequencing reactions were conducted at GENEWIZ, Inc. (South Plainfield, NJ, USA), and the NEBNext Ultra DNA library preparation kit (New England Biolabs) was used following the manufacturer's recommendations. The DNA library was quantified by real-time PCR (Applied Biosystems, Carlsbad, CA, USA) and was loaded onto a HiSeq instrument according to the manufacturer's instructions (Illumina). Sequencing was performed using a 2 × 150-bp paired-end (PE) configuration in high-output run mode; image analysis and base calling were conducted by the HiSeq Control Software (HCS) on the HiSeq instrument. The .bcl files were converted to fastq files and demultiplexed using Illumina bcl2fastq software. Sequence reads were trimmed with Trimmomatic v.0.38, using the following options: ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36. Trimmed reads were assembled using MEGAHIT v.1.2.9 with default parameters together with the default preset meta-large options to produce a single coassembly. We obtained 250,051,961 × 2-bp PE reads for EFPC-H and 200,408,672 × 2-bp PE reads for EFPC-F samples. Reads were mapped to the assembly with Bowtie 2 (4) and were used for binning into MAGs using MetaBAT 2 (5). A total of 28 MAGs of medium or higher quality (i.e., completeness of ≥50% and contamination of ≤10%) (6), based on CheckM results (7), were selected for the announcement. Notably, 27 of the MAGs represent novel species as defined by the Genome Taxonomy Database (GTDB) (release R05-RS95) (8). A more detailed list of the assembly statistics and taxonomy of all MAGs can be found in Table 1. Furthermore, the relative abundances of the MAGs in the

Citation Kim M, Wilpiszkeski RL, Wells M, Wymore AM, Gionfriddo CM, Brooks SC, Podar M, Elias DA. 2021. Metagenome-assembled genome sequences of novel prokaryotic species from the mercury-contaminated East Fork Poplar Creek, Oak Ridge, Tennessee, USA. *Microbiol Resour Announc* 10:e00153-21. <https://doi.org/10.1128/MRA.00153-21>.

Editor Frank J. Stewart, Georgia Institute of Technology

This is a work of the U.S. Government and is not subject to copyright protection in the United States. Foreign copyrights may apply. Address correspondence to Dwayne A. Elias, eliasda@ornl.gov.

Received 18 February 2021

Accepted 31 March 2021

Published 29 April 2021

TABLE 1 MAG statistics, taxonomy, and relative abundances in metagenomes

Identification no.	Taxonomy	GenBank accession no.	Total length (bp)	Completeness (%)	Contamination (%)	Relative abundance in EFPC-H (%)	Relative abundance in EFPC-K (%)
1	Domain <i>Bacteria</i> , phylum <i>Acidobacteriota</i> , class <i>Aminicenantia</i> , order <i>Aminicenantiales</i> , family <i>RBG-16-66-30</i> , genus <i>RBG-16-66-30</i>	JAEHCW0000000000	2,374,023	56.06	6.64	0	0.134
116	Domain <i>Bacteria</i> , phylum <i>Acidobacteriota</i> , class <i>Aminicenantia</i> , order <i>UBA2199</i> , family <i>UBA2199</i>	JAEHCX0000000000	2,093,061	56.05	2.78	0	0.082
118	Domain <i>Bacteria</i> , phylum <i>Proteobacteria</i> , class <i>Gammaproteobacteria</i> , order <i>Chromatiales</i> , family <i>Chromatiaceae</i>	JAEHCY0000000000	3,039,835	88.78	3.36	0.005	0.095
12	Domain <i>Archaea</i> , phylum <i>Thermoproteota</i> , class <i>Nitrososphaeria</i> , order <i>Nitrososphaerales</i> , family <i>Nitrososphaeraceae</i> , genus <i>TA-21</i>	JAEHCZ0000000000	1,598,264	64.51	6.15	0.105	0.012
121	Domain <i>Bacteria</i> , phylum <i>Proteobacteria</i> , class <i>Gammaproteobacteria</i> , order <i>Burkholderiales</i> , family <i>Rhodocyclaceae</i> , genus <i>Fen-999</i>	JAEHDA0000000000	2,560,960	90.95	6.02	0.001	0.205
123	Domain <i>Bacteria</i> , phylum <i>Proteobacteria</i> , class <i>Gammaproteobacteria</i> , order <i>Burkholderiales</i> , family <i>Palsa-1005</i> , genus <i>VBCG01</i>	JAEHDB0000000000	3,834,344	86.84	7.71	0.021	0.074
124	Domain <i>Bacteria</i> , phylum <i>Patescibacteria</i> , class <i>Saccharimonadales</i> , order <i>Saccharimonadales</i>	JAEHDC0000000000	954,763	57.36	0	0	0.082
133	Domain <i>Bacteria</i> , phylum <i>Proteobacteria</i> , class <i>Alphaproteobacteria</i> , order <i>Rhizobiales</i> , family <i>Hyphomicrobiaceae</i> , genus <i>AWTP1-13</i>	JAEHDD0000000000	4,018,257	67.4	7.64	0.114	0.019
134	Domain <i>Bacteria</i> , phylum <i>Actinobacteriota</i> , class <i>Acidimicrobia</i> , order <i>Acidimicrobiales</i>	JAEHDE0000000000	4,891,318	78.91	6.41	0.078	0.005
137	Domain <i>Archaea</i> , phylum <i>Thermoproteota</i> , class <i>Nitrososphaeria</i> , order <i>Nitrososphaerales</i> , family <i>Nitrosopumilaceae</i> , genus <i>Nitrosarchaeum</i> , species <i>Nitrosarchaeum sp005798405</i>	JAEHDF0000000000	827,553	59.01	2.27	0.021	0.055
139	Domain <i>Bacteria</i> , phylum <i>Proteobacteria</i> , class <i>Gammaproteobacteria</i> , order <i>UBA6522</i> , family <i>UBA6522</i> , genus <i>FEN-1219</i>	JAEH DG0000000000	2,723,786	58.18	5.12	0.107	0.005
15	Domain <i>Bacteria</i> , phylum <i>Spirochaetota</i> , class <i>GWEE2-31-10</i> , order <i>GWEE2-31-10</i> , family <i>GWEE2-31-10</i>	JAEH DH0000000000	3,566,009	76.98	5.2	0	0.056

(Continued on next page)

TABLE 1 (Continued)

Identification no.	Taxonomy	GenBank accession no.	Total length (bp)	Completeness (%)	Contamination (%)	Relative abundance in EFPC-H (%)	Relative abundance in EFPC-K (%)
158	Domain <i>Bacteria</i> , phylum <i>Acidobacteriota</i> , class <i>Acidobacteriae</i> , order <i>Bryobacteriales</i> , family <i>Bryobacteraceae</i>	JAEHDI0000000000	5,740,825	90.35	6.48	0.055	0.041
17	Domain <i>Bacteria</i> , phylum <i>Myxococota</i> , class <i>Myxococcia</i> , order <i>Myxococcales</i> , family <i>Anaeromyxobacteraceae</i> , genus <i>Anaeromyxobacter</i>	JAEHDI0000000000	2,526,394	50.96	3.82	0.001	0.085
2	Domain <i>Bacteria</i> , phylum <i>Actinobacteriota</i> , class <i>Actinomycetia</i> , order <i>Mycobacteriales</i> , family <i>Micromonosporaceae</i> , genus <i>SLSW01</i>	JAEHDK0000000000	5,523,212	90.19	2.88	0.106	0.002
24	Domain <i>Bacteria</i> , phylum <i>Methyloirabiolota</i> , class <i>Methyloirabiolia</i> , order <i>Rokubacteriales</i> , family <i>GWA2-73-35</i> , genus <i>AR5</i>	JAEHDL0000000000	4,673,507	68.97	5.78	0.149	0.015
41	Domain <i>Bacteria</i> , phylum <i>Actinobacteriota</i>	JAEHDM0000000000	3,219,401	62.59	3.59	0.097	0.006
51	Domain <i>Bacteria</i> , phylum <i>Eisenbacteria</i> , class <i>RBG-16-71-46</i> , order <i>SZUA-252</i> , family <i>SZUA-252</i> , genus <i>WS-9</i>	JAEHDN0000000000	2,493,152	78.32	0.55	0.111	0.008
67	Domain <i>Bacteria</i> , phylum <i>Proteobacteria</i> , class <i>Gammaproteobacteria</i> , order <i>Burkholderiales</i> , family <i>Rhodocyclaceae</i> , genus <i>UTPRO2</i>	JAEHDO0000000000	2,399,743	84.91	2.43	0.024	0.423
68	Domain <i>Bacteria</i> , phylum <i>Proteobacteria</i> , class <i>Gammaproteobacteria</i> , order <i>Beggiatoales</i> , family <i>Beggiatoaceae</i> , genus <i>Thioploca</i>	JAEHDP0000000000	2,206,680	56.12	1.31	0	0.088
69	Domain <i>Bacteria</i> , phylum <i>Proteobacteria</i> , class <i>Gammaproteobacteria</i> , order <i>Pseudomonadales</i> , family <i>Haliaceae</i> , genus <i>Halioglobus</i>	JAEHQ0000000000	2,575,066	71.32	7.01	0.056	0.036
7	Domain <i>Bacteria</i> , phylum <i>Acidobacteriota</i> , class <i>Thermoanaerobaculia</i> , order <i>UBA5066</i> , family <i>Gp7-AA6</i>	JAEHDR0000000000	2,531,270	53.45	5.17	0.058	0.007
70	Domain <i>Bacteria</i> , phylum <i>Acidobacteriota</i> , class <i>Thermoanaerobaculia</i> , order <i>UBA5066</i> , family <i>Gp7-AA6</i> , genus <i>Gp7-AA6</i>	JAEHDS0000000000	2,820,809	76.55	5.24	0.057	0.014
76	Domain <i>Bacteria</i> , phylum <i>Gemmatimonadota</i> , class <i>Gemmatimonadetes</i> , order <i>Gemmatimonadales</i> , family <i>GW2-71-9</i> , genus <i>AG41</i>	JAEHDT0000000000	2,174,421	68.32	3.05	0.151	0.012

(Continued on next page)

TABLE 1 (Continued)

Identification no.	Taxonomy	GenBank accession no.	Total length (bp)	Completeness (%)	Contamination (%)	Relative abundance in EFPC-H (%)	Relative abundance in EFPC-K (%)
8	Domain <i>Bacteria</i> , phylum <i>Gemmatimonadota</i> , class <i>Gemmatimonadetes</i> , order <i>Gemmatimonadales</i> , family <i>GWC2-71-9</i>	JAEHDX0000000000	2,995,922	69.94	6.52	0.252	0.01
84	Domain <i>Bacteria</i> , phylum <i>Nitrospirota</i> , class <i>Nitrospiria</i> , order <i>Nitrospirales</i> , family <i>Nitrospiraceae</i> , genus <i>Palsa-1315</i>	JAEHDX0000000000	2,466,425	82.94	3.36	0.1	0.184
85	Domain <i>Bacteria</i> , phylum <i>Acidobacteriota</i> , class <i>Vicinamibacteria</i> , order <i>Vicinamibacteriales</i> , family <i>UBA2999</i>	JAEHDX0000000000	3,440,127	68	6.44	0.09	0.014
95	Domain <i>Bacteria</i> , phylum <i>Proteobacteria</i> , class <i>Gammaproteobacteria</i> , order <i>GCA-2729495</i> , family <i>GCA-2729495</i>	JAEHDX0000000000	2,267,503	51.21	3.45	0.007	0.069

two metagenomes were calculated by competitive Bowtie 2 alignment (4) and then normalized for genome equivalents in the metagenomic data size with averaged genome size for the microbial community samples using MicrobeCensus (9). Additionally, we calculated the 80% truncated average depth (TAD80) for MAG abundance using the BedGraph.tad.rb script of the enveomics collection (10) to remove outlier genes in terms of coverage, such as the rRNA genes and other multicopy genes. All MAGs represented 0% to 0.4% of the total microbial community of either EFPC-H or EFPC-K (Table 1). Each of the 28 MAGs was annotated using the RASTtk genome annotation program (11) made available through the PATRIC Bioinformatics Resource Center (12). Consistent with the history of anthropogenic contamination with mercury at the EFPC, 27 of the 28 MAGs (all except MAG 124, a *Saccharimonadales* bacterium from the *Pastescibacteria* phylum) contained either putative heavy metal resistance genes or ATPase efflux pump genes. Notably, 17 of the 28 MAGs contained traits for Se-assimilatory metabolism, including the candidate phyla *Eisenbacteria* and *Rokubacteria*, which marks the first report of Se metabolism in these bacterial candidate phyla. Default parameters were used for all software unless otherwise noted.

Data availability. Raw sequence reads are available under BioProject accession number [PRJNA670906](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA670906). Genome sequences of MAGs are available under the GenBank accession numbers provided in Table 1.

ACKNOWLEDGMENTS

This research was sponsored in part by the Office of Biological and Environmental Research, Office of Science, U.S. Department of Energy (DOE), as part of the Mercury Science Focus Area at Oak Ridge National Laboratory, which is managed by UT-Battelle LLC for the DOE under contract DE-AC05-00OR22725. This manuscript has been authored by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy.

REFERENCES

- Brooks SC, Southworth GR. 2011. History of mercury use and environmental contamination at the Oak Ridge Y-12 plant. *Environ Pollut* 159:219–228. <https://doi.org/10.1016/j.envpol.2010.09.009>.
- Vishnivetskaya TA, Mosher JJ, Palumbo AV, Yang ZK, Podar M, Brown SD, Brooks SC, Gu B, Southworth GR, Drake MM, Brandt CC, Elias DA. 2011. Mercury and other heavy metals influence bacterial community structure in contaminated Tennessee streams. *Appl Environ Microbiol* 77:302–311. <https://doi.org/10.1128/AEM.01715-10>.
- Christensen GA, Gionfriddo CM, King AJ, Moberly JG, Miller CL, Somenahally AC, Callister SJ, Brewer H, Podar M, Brown SD, Palumbo AV, Brandt CC, Wymore AM, Brooks SC, Hwang C, Fields MW, Wall JD, Gilmour CC, Elias DA. 2019. Determining the reliability of measuring mercury cycling gene abundance with correlations with mercury and methylmercury concentrations. *Environ Sci Technol* 53:8649–8663. <https://doi.org/10.1021/acs.est.8b06389>.
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>.
- 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>.
- Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, Schulz F, Jarett J, Rivers AR, Eloe-Fadrosh EA, Tringe SG, Ivanova NN, Copeland A, Clum A, Becraft ED, Malmstrom RR, Birren B, Podar M, Bork P, Weinstock GM, Garrity GM, Dodsworth JA, Yooshep S, Sutton G, Glöckner FO, Gilbert JA, Nelson WC, Hallam SJ, Jungbluth SP, Ettema TJG, Tighe S, Konstantinidis KT, Liu W-T, Baker BJ, Rattai T, Eisen JA, Hedlund B, McMahon KD, Fierer N, Knight R, Finn R, Cochrane G, Karsch-Mizrachi I, Tyson GW, Rinke C, Genome Standards Consortium, Lapidus A, Meyer F, Yilmaz P, Parks DH, Eren AM, Schriml L, Banfield JF, Hugenholtz P, Woyke T. 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>.
- 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>.
- Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2020. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* 36:1925–1927. <https://doi.org/10.1093/bioinformatics/btz848>.
- Nayfach S, Pollard KS. 2015. Average genome size estimation improves comparative metagenomics and sheds light on the functional ecology of the human microbiome. *Genome Biol* 16:51. <https://doi.org/10.1186/s13059-015-0611-7>.
- Rodriguez-R LM, Konstantinidis KT. 2016. The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. *PeerJ Preprints* 4:e1900v1. <https://doi.org/10.7287/peerj.preprints.1900v1>.
- Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Thomason JA, Stevens R, Vonstein V, Wattam AR, Xia F. 2015. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci Rep* 5:8365. <https://doi.org/10.1038/srep08365>.
- Davis JJ, Wattam AR, Aziz RK, Brettin T, Butler R, Butler RM, Chlenski P, Conrad N, Dickerman A, Dietrich EM, Gabbard JL, Gerdes S, Guard A, Kenyon RW, Machi D, Mao C, Murphy-Olson D, Nguyen M, Nordberg EK, Olsen GJ, Olson RD, Overbeek JC, Overbeek R, Parrello B, Pusch GD, Shukla M, Thomas C, VanOeffelen M, Vonstein V, Warren AS, Xia F, Xie D, Yoo H, Stevens R. 2020. The PATRIC Bioinformatics Resource Center: expanding data and analysis capabilities. *Nucleic Acids Res* 48:D606–D612. <https://doi.org/10.1093/nar/gkz943>.