Five Metagenome-Assembled Genomes of the Rare Phylum CSSED10-310 from Zodletone Spring (Oklahoma, USA)

Archana Yadav,[a] C. Ryan Hahn,[a] Mostafa S. Elshahed,[a] (ID) Noha H. Youssef[a]

[a]Department of Microbiology and Molecular Genetics, Oklahoma State University, Stillwater, Oklahoma, USA

ABSTRACT We analyzed five metagenome-assembled genomes (MAGs) belonging to the rare, yet-uncultured phylum CSSED10-310 recovered from the anoxic sediments of Zodletone Spring (Oklahoma). Our analysis suggests their potential involvement in sulfite respiration.

Zodletone Spring is a surficial, anoxic, sulfide- and sulfur-rich spring in southwestern Oklahoma. Prior studies have documented the phylogenetic diversity in the spring (1–5). Such studies have demonstrated that the spring harbors a plethora of novel and rare taxa. Here, we report on the assembly and analysis of five genomes belonging to the rare, yet-uncultured phylum CSSED10-310. Currently (April 2021), this phylum is represented in the Genome Taxonomy Database (GTDB; release 95) by a single genome (GCA_003558985.1) binned from sediments of a hypersaline soda lake (6). The phylum appears to be a sister phylum to the *Acidobacteriota*.

Samples from the anoxic, sulfide-saturated source sediments were obtained from Zodletone Spring in September 2017. Ten samples were collected from 5 cm deep into the anoxic sediments by completely filling sterile 50-ml polypropylene plastic tubes. The tubes were kept on ice until they were brought back to the lab (~2-h drive), where they were immediately processed. DNA extraction was conducted on 0.5 g sediment from each of the 10 replicate samples using the DNeasy PowerSoil kit (Qiagen, Valencia, CA, USA) according to manufacturer's protocols. All DNA extractions were pooled and used for the preparation of sequencing libraries using the Nextera XT DNA library prep kit (Illumina, San Diego, CA, USA) as per the manufacturer's instructions. Sequencing was conducted using the Illumina HiSeq 2500 platform using the services of Novogene (Beijing, China), generating 281 Gbp of 150-bp paired-end raw sequence output. FastQC v0.11.5 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was used to assess the quality of the reads, followed by trimming using Trimmomatic v0.38 (7). High-quality reads were assembled into contigs using MEGAHIT v1.1.3 (8). MetaBAT 2 v1.7 (9) and MaxBin 2 v2.2.4 (10) were used to bin the contigs into draft genomes, and DasTool v1.1.1-0 (11) was used to select the highest-quality bins. Genome completeness, strain heterogeneity, and contamination were estimated using CheckM v1.1.3 (12). Default parameters were used except where otherwise noted. GhostKOALA (13) was used for functional annotation by assigning protein-coding genes to KEGG orthologies (KOs). KEGG mapper (14) was used to visualize metabolic pathways for this phylum. The taxonomic affiliation of the genomes was determined using GTDB-Tk v1.1.0 (15, 16), and the generated concatenated alignment was used to construct a maximum likelihood phylogenomic tree using FastTree (17).

Five genomes recovered from the spring source sediment metagenome were affiliated with the rare, yet-uncultured phylum CSSED10-310 (Fig. 1). Sequencing statistics (including the number of contigs, median genome coverage, and N_{50} value) and general genomic features of the CSSED10-310 genomes are shown in Table 1. The expected genome sizes ranged from 3.01 to 5.72 Mbp, and the GC content ranged from 43.4 to 58.9%. The cells are predicted to be Gram negative and possibly motile (based on the identification of the
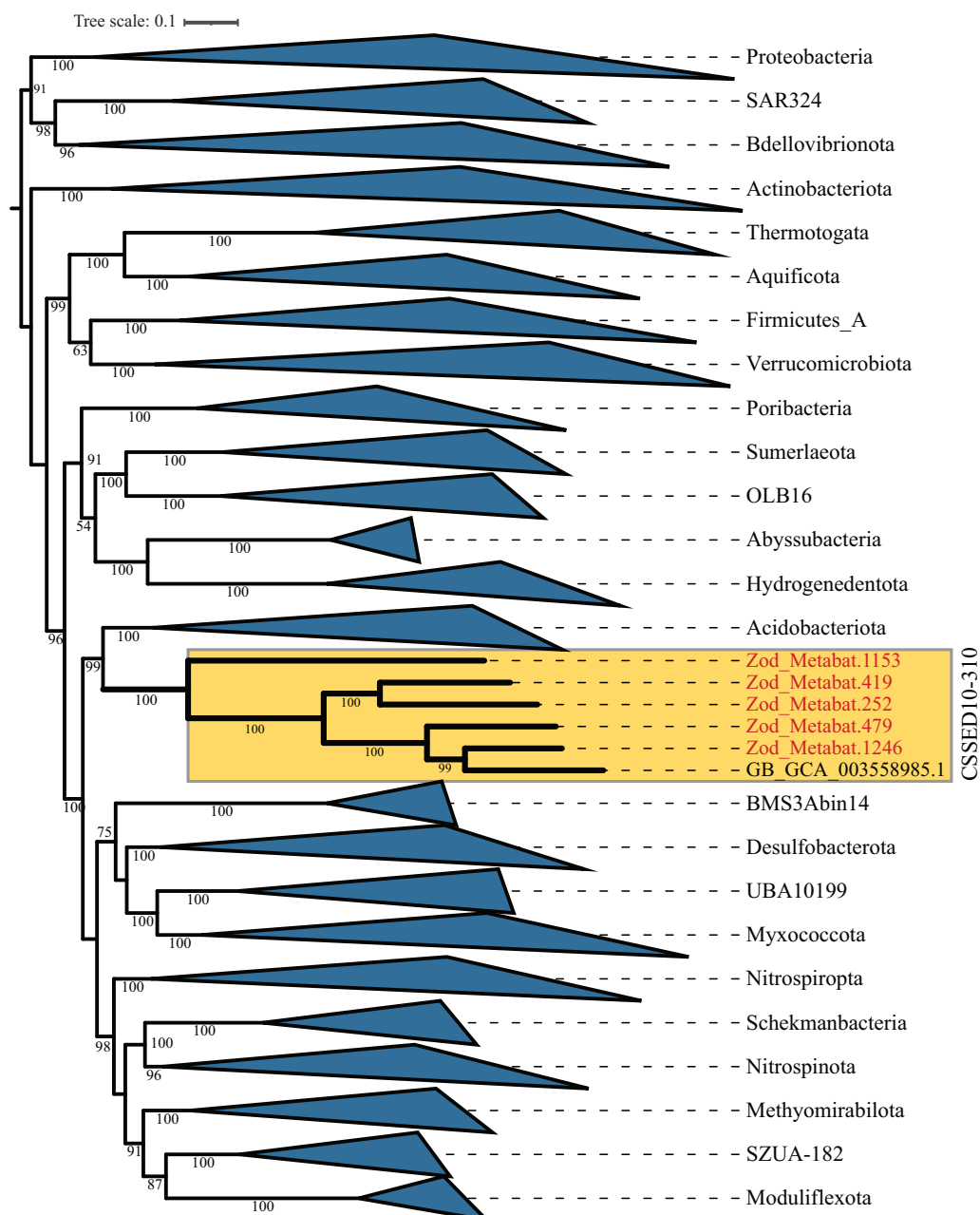
FIG 1 Maximum likelihood tree based on the concatenated alignment of 120 single-copy marker genes showing the phylogenetic position of phylum CSSED10-310 relative to other phyla. The tree was constructed in FastTree (17) and visualized using iTOL (18). Phylum CSSED10-310 is highlighted in yellow, and all other phyla are wedged. The 5 MAGs from Zodletone Spring discussed here are shown in red bold text. Names depict the MAG bin name (as shown in Table 1). The single CSSED10-310 genome (assembly accession number GCA_003558985.1) available in GTDB is also highlighted in the same clade. The tree was midpoint rooted, and the bootstrap values (from 100) are displayed for the branches with ≥50% support.

majority of flagellum and type IV pilus biosynthesis and assembly genes). A heterotrophic lifestyle is predicted, with sugars (glucose, fructose, mannose, ribulose, and galactose), starch, and propionate as potential carbon sources. Two genomes (Zod_Metabat.252 and Zod_Metabat.419) encoded the anaerobic sulfite reductase (AsrABC) system, as well as the membrane-bound heterodisulfide reductase-related enzymes (HdrABC) for transfer of electrons to the AsrC subunit, suggesting sulfite reduction capacities coupled to sugar degradation as an energy-generating process in the analyzed phylum CSSED10-310 genomes. In addition, the genomes encoded sugar fermentative capabilities.

TABLE 1 General genomic features of the five MAGs studied

Parameter	Data for indicated MAG bin name				
	Zod_Metabat.1153	Zod_Metabat.1246	Zod_Metabat.252	Zod_Metabat.419	Zod_Metabat.479
GenBank assembly accession no.	JAFGDC000000000	JAFGEQ000000000	JAFGJC000000000	JAFGLW000000000	JAFGMN000000000
Completeness (%)	86.32	85.42	68.11	82.85	85.59
Contamination (%)	4.75	0.85	0	1.28	5.13
GTDB classification					
Phylum	CSSED10-310	CSSED10-310	CSSED10-310	CSSED10-310	CSSED10-310
Class	CSSED10-310	CSSED10-310	CSSED10-310	CSSED10-310	CSSED10-310
Order	Novel order ZNO13[a]	CSSED10-310	CSSED10-310	CSSED10-310	CSSED10-310
Family	Novel family ZNF029[b]	CSSED10-310	Novel family ZNF028[b]	Novel family ZNF028[b]	CSSED10-310
No. of contigs	635	24	522	31	360
Assembly N_{50} (bp)	9,603	245,466	9,322	180,367	13,125
Median genome coverage ($\times$)	9.54	30.39	8.98	25.83	9.59
Assembly size (Mbp)	4.94	2.57	3.85	3.12	3.46
Expected genome size (Mbp)	5.72	3.01	5.65	3.76	4.04
GC content (%)	58.9	48.9	43.4	54	53.2
No. of genes	4,044	2,073	3,368	2,515	2,993
Coding %	89.05	94.16	93.29	93.96	94.09
Avg gene length (bp)	1,087	1,168	1,066	1,165	1,087
No. of rRNA genes					
5S rRNA	1	1	0	1	1
16S rRNA	0	0	0	0	0
23S rRNA	0	0	0	0	0
No. of tRNA genes	34	35	18	41	30
Quality tiers	MQD[c]	MQD	MQD	MQD	MQD
CRISPR count	9	3	1	6	1

[a] Genomes were unclassified by the GTDB-Tk at the order level and were assigned to a novel order.
[b] Genomes were unclassified by the GTDB-Tk at the family level and were assigned to a novel family.
[c] MQD, medium quality draft.

Data availability. Raw sequencing reads were deposited in the SRA under accession number SRX9813571. The whole-genome shotgun project was submitted to GenBank under BioProject number PRJNA690107 and BioSample number SAMN17269717. The individual metagenome-assembled genomes (MAGs) have been deposited at DDBJ/ENA/GenBank under the accession numbers JAFGEQ000000000, JAFGDC000000000, JAFGJC000000000, JAFGMN000000000, and JAFGLW000000000 and were annotated using the NCBI Prokaryotic Genome Annotation Pipeline.

REFERENCES

1. Elshahed MS, Senko JM, Najar FZ, Kenton SM, Roe BA, Dewers TA, Spear JR, Krumholz LR. 2003. Bacterial diversity and sulfur cycling in a mesophilic sulfide-rich spring. Appl Environ Microbiol 69:5609–5621. https://doi.org/10.1128/AEM.69.9.5609-5621.2003.
2. Spain AM, Elshahed MS, Najar FZ, Krumholz LR. 2015. Metatranscriptomic analysis of a high-sulfide aquatic spring reveals insights into sulfur cycling and unexpected aerobic metabolism. Peer J 3:e1259. https://doi.org/10.7717/peerj.1259.
3. Youssef NH, Couger MB, Elshahed MS. 2010. Fine-scale bacterial beta diversity within a complex ecosystem (Zodletone spring, OK, USA): the role of the rare biosphere. PLoS One 5:e12414. https://doi.org/10.1371/journal.pone.0012414.
4. Youssef NH, Farag IF, Hahn CR, Premathilake H, Fry E, Hart M, Huffaker K, Bird E, Hambright J, Hoff WD, Elshahed MS. 2019. Candidatus Krumholzibacterium zodletonense gen. nov., sp nov., the first representative of the candidate phylum Krumholzibacteriota phyl. nov. recovered from an anoxic sulfidic spring using genome resolved metagenomics. Syst Appl Microbiol 42:85–93. https://doi.org/10.1016/j.syapm.2018.11.002.
5. Elshahed MS, Najar FZ, Aycock M, Qu C, Roe BA, Krumholz LR. 2005. Metagenomic analysis of the microbial community at Zodletone Spring (Oklahoma): insights into the genome of a member of the novel candidate division OD1.
Appl Environ Microbiol 71:7598–7962. https://doi.org/10.1128/AEM.71.11.7598-7602.2005.
6. Vavourakis CD, Andrei A-S, Mehrshad M, Ghai R, Sorokin DY, Muyzer G. 2018. A metagenomics roadmap to the uncultured genome diversity in hypersaline soda lake sediments. Microbiome 6:168. https://doi.org/10.1186/s40168-018-0548-7.
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. Li D, Liu CM, Luo R, Sadakane K, Lam TW. 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.
9. 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.
10. Wu YW, Simmons BA, Singer SW. 2016. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics 32:605–607. https://doi.org/10.1093/bioinformatics/btv638.

11. Sieber CMK, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, Banfield JF. 2018. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. Nat Microbiol 3:836–843. https://doi.org/10.1038/s41564-018-0171-1.

12. 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.

13. 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.

14. Kanehisa M. 2019. Toward understanding the origin and evolution of cellular organisms. Protein Sci 28:1947–1951. https://doi.org/10.1002/pro.3715.

15. Parks DH, Chuvochina M, Chaumeil PA, Rinke C, Mussig AJ, Hugenholtz P. 2020. A complete domain-to-species taxonomy for Bacteria and Archaea. Nat Biotechnol 38:1079–1086. https://doi.org/10.1038/s41587-020-0501-8.

16. Chaumeil PA, 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.

17. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2: approximately maximum-likelihood trees for large alignments. PLoS One 5:e9490. https://doi.org/10.1371/journal.pone.0009490.

18. Letunic I, Bork P. 2019. Interactive Tree of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res 47:W256–W259. https://doi.org/10.1093/nar/gkz239.