Archana Yadav[a], C. Ryan Hahn[a], Mostafa S. Elshahed[a], and Noha H. Youssef[a#]

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

Running Head: Phylum CSSED10-310 genomes from an anoxic spring.

#Address correspondence to Noha H. Youssef (Noha@Okstate.edu).

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, USA). Our analysis suggests their potential involvement in sulfite respiration.

Announcement

Zodletone spring is a surficial, anoxic, sulfide and sulfur-rich spring in Southwestern Oklahoma, USA. 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 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 5-cm deep into the anoxic sediments by completely filling sterile 50-mL polypropylene plastic tubes. Tubes were kept on ice until brought back to the lab (~2h drive), where they were immediately processed. DNA extraction was conducted on 0.5 g sediment from each of the ten replicate samples using the DNeasy PowerSoil kit (Qiagen, Valencia, CA, USA) according to manufacturer 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 manufacturers' instructions. Sequencing was conducted using the Illumina HiSeq 2500 platform using the services of Novogene (Beijing, China), generating 281 Gbp of 150-bp pair-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 (v.1.1.3) (8). MetaBAT2 v1.7 (9) and MaxBin2 - v2.2.4 (10) were used to bin the contigs into draft genomes, and DasTool v 1.1.1-0 (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 were 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 (Figure 1). Sequencing statistics (including number of contigs, median genome coverage and N50) as well as general genomic features of CSSED10-310 genomes are shown in Table 1. Expected genome sizes ranged from 3.01-5.72 Mbp, and GC content ranged from 43.4-58.9%. Cells are predicted to be Gram-negative, and possibly motile (based on the identification of the majority of flagella and type IV pili 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 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.

Data Availability. Raw sequencing reads were deposited in the SRA under accession SRX9813571. The whole genome shotgun project was submitted to GenBank under Bioproject ID PRJNA690107 and Biosample ID SAMN17269717. The individual assembled MAGs have been deposited at DDBJ/ENA/GenBank under the accession JAFGEQ000000000, JAFGDC000000000, JAFGJC000000000, JAFGMN000000000 and JAFGLW000000000, and were annotated using NCBI Prokaryotic Genome Annotation Pipeline.

Figure 1 Legend

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 GB_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.

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.
2. Spain AM, Najar FZ, Krumholz LR, Elshahed MS. 2015. Metatranscriptomic analysis of a high-sulfide aquatic spring reveals insights into sulfur cycling and unexpected aerobic metabolism. Peer J 3:e1259.
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.
4. Youssef NH, Farag IF, Hahn CR, Prematthilake 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.
5. Elshahed MS, Najar FZ, Aycock M, Qu C, Roe BA, Krumholz LR. 2005. Metagenomci analysis of the microbial community at Zodletone spring: Insights into the genome of novel candidate division OD1. Appl Environ Microbiol 71:7958-7962.
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.
7. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114-20.
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 31:1674-6.
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.
10. Wu YW, Simmons BA, Singer SW. 2016. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics 32:605-7.
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.
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 25:1043-55.
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.
14. Kanehisa M. 2019. Toward understanding the origin and evolution of cellular organisms. Protein Sci 28:1947-1951.

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.
16. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. 2019. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics doi: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.
18. Letunic I, Bork P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Research 47:W256-W259.