Metagenomic Sequencing of a Patescibacteria-containing enrichment from Zodletone Spring in Oklahoma, USA

Sayali A. Mulay[a], C. Ryan Hahn[b], Dawn M. Klingeman[c], Mostafa S. Elshahed[b], Noha H. Youssef[b] and Mircea Podar[a,c*]

[a]Department of Microbiology, University of Tennessee Knoxville, Knoxville, Tennessee, USA
[b]Department of Microbiology and Molecular Genetics, Oklahoma State University, Stillwater, Oklahoma, USA
[c]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA

Running head: Metagenome of a Zodletone Spring enrichment

*Address correspondence to Mircea Podar at podarm@ornl.gov

Abstract

An enrichment of sulfidic sediments from Zodletone spring was sequenced as a metagenome. Draft genomes representing Cloacimonadota, Deltabacterota, Firmicutes, and Patescibacteria were binned, annotated, and will aid functional genomics and cultivation efforts.

Zodletone spring is a sulfidic spring in Kiowa County, Oklahoma (GPS coordinates 35.002444 N, 98.688167 W), used as a model environment to study microbial physiological adaptations to redox gradients resembling the great oxygenation event of early Earth (1). Among the over 60 archaeal and bacterial high-level taxa that inhabit Zodletone, Patescibacteria are highly represented and include more than two dozen largely uncultured lineages originally assigned under the bacterial "Candidate Phyla Radiation" (2-4). Zodletone sediment samples (50 ml) were collected as described (2) and were maintained anoxic and refrigerated in glass bottles with nitrogen headspace. Enrichments were generated by inoculating ~1 ml sediment into 10 ml of Pfennig's mineral medium containing 0.12% yeast extract, 8mM sodium sulfite, reduced with 5mM cysteine under 85% N_2, 10% CO_2 and 5% H_2, followed by incubation at 25^{o}C and weekly passaged (1:10) in fresh medium. We obtained a Zodletone sediment anaerobic enrichment dominated by Deltabacterota (>50%, based on SSU rRNA amplicons, performed as described in (2)), including a *Desulfomicrobium* we have already isolated in pure culture and sequenced to completion (5) as well as *Desulfovibrio, Desulfuromonas* and *Desulfocapsa*. The enrichment contained Patescibacteria, represented by Absconditabacterales (formerly SR1, candidate class Gracilibacteria) and Moranbacterales (candidate class Paceibacteria) at relatively low levels (<5%), as well as Spirochaetota (*Sphaerochaeta*), Cloacimonadota and Firmicutes.

For metagenomic DNA isolation, the enrichment was grown for 7 days at 25^{o}C in 10 ml medium. All experimental protocols followed manufacturers' instructions. DNA was isolated

using the ZymoBIOMICS DNA Miniprep Kit (Zymo Research, Irvine CA). A short insert library was prepared using the Illumina Nextera XT DNA Library Preparation Kit. Sequencing (Illumina MiSeq Reagent Kit v2, 2x250-nucleotide reads) on a MiSeq instrument (Illumina, Inc., San Diego, CA), yielded 11.6 million paired reads (236 nt average length). The sequences were imported in KBase (6), and analyzed following the KBase standard metagenomic workflow (7) using default settings unless otherwise noted. Trimmomatic v0.36 (8) was used for quality-based trimming and to remove reads shorter than 150 nt. The reads were assembled using metaSPAdes v3.15.3 (9), MEGAHIT v1.2.9 (10) and IDBA-UD v1.1.3 (11) and the assemblies statistics compared with QUAST v4.4 (12). The metaSPADES assembly was selected for binning based on highest scores. The best output in terms of genome completion and contamination score, as assessed by CheckM v1.0.18 (13), was generated by MetaBAT2 v1.7 (14), which generated 21 bins (metagenome-assembled genomes, MAGs). The MAGs were taxonomically classified with GTDB-Tk v2.3.2 (15). The most complete (85-97%) and least contaminated (<2.5%), based on single copy gene statistics, belong to Desulfobacterota (*Desulfolutivibrio, Pseudodesulfovibrio, Desulfovibrio*), Cloacimonadota and Acholeplasmatales. Less complete MAGs (<70%) represented *Sphaerochaeta, Desulfocapsa, Desulfomicrobium* and four MAGs of Bacillota. Our primary target group, Patescibacteria, included MAGs for an Absconditabacterales and a Moranbacterales (65-85% complete, <1% contamination). For microbes with reduced genomes such as Patescibacteria (16), completion based on universal single copy genes is underestimated. For example, the closed genome of *Cand.* Absconditicoccus praedator, an Absconditabacterales, (17) is predicted as 80% complete by CheckM.

To predict and annotate the genes in the high-quality MAGs, we used the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) v6.1 (18). Table 1 lists the genome-based classification and MAGs/gene content information. The genomes and their functional annotations will facilitate mass-spectrometry proteomics and metabolic predictions to aid cultivation of individual bacteria from that enrichment using selective media and targeted reverse genomics (19).

Data availability. The metagenome sequence reads are available in the NCBI Sequence Read Archive (SRA) under the accession number SRR27793275. The annotated MAGs have been deposited in GenBank under the accession numbers JAZHDB000000000, JAZHDC000000000, JAZHDD000000000, JAZHDE000000000, JAZHDF000000000, JAZHDG000000000 and JAZHDH000000000.

Acknowledgements. This research was funded by the National Science Foundation grants 2016371 (to M.P.) and 2016423 (to N.H.Y. and M.S.E.). Oak Ridge National Laboratory is managed by UT-Battelle, LLC., for the U.S. Department of Energy under contract DE-AC05-00OR22725.

1. Buhring SI, Sievert SM, Jonkers HM, Ertefai T, Elshahed MS, Krumholz LR, Hinrichs KU. 2011. Insights into chemotaxonomic composition and carbon cycling of phototrophic communities in an artesian sulfur-rich spring (Zodletone, Oklahoma, USA), a possible analog for ancient microbial mat systems. Geobiology 9:166-79.
2. Hahn CR, Farag IF, Murphy CL, Podar M, Elshahed MS, Youssef NH. 2022. Microbial Diversity and Sulfur Cycling in an Early Earth Analogue: From Ancient Novelty to Modern Commonality. mBio 13:e0001622.

3. Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN, Hernsdorf AW, Amano Y, Ise K, Suzuki Y, Dudek N, Relman DA, Finstad KM, Amundson R, Thomas BC, Banfield JF. 2016. A new view of the tree of life. Nat Microbiol 1:16048.
4. Parks DH, Rinke C, Chuvochina M, Chaumeil PA, Woodcroft BJ, Evans PN, Hugenholtz P, Tyson GW. 2017. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. Nat Microbiol 2:1533-1542.
5. Mulay SA, Alexander WG, Hahn CR, Klingeman DM, Elshahed MS, Youssef NH, Podar M. 2023. Complete Genome Sequence of Desulfomicrobium sp. Strain ZS1 from Zodletone Spring in Oklahoma, USA. Microbiol Resour Announc 12:e0014523.
6. 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 JM, Chia JM, 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.
7. Chivian D, Jungbluth SP, Dehal PS, Wood-Charlson EM, Canon RS, Allen BH, Clark MM, Gu T, Land ML, Price GA, Riehl WJ, Sneddon MW, Sutormin R, Zhang Q, Cottingham RW, Henry CS, Arkin AP. 2022. Metagenome-assembled genome extraction and analysis from microbiomes using KBase. Nat Protoc doi:10.1038/s41596-022-00747-x.
8. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114-20.
9. Antipov D, Korobeynikov A, McLean JS, Pevzner PA. 2015. hybridSPAdes: an algorithm for hybrid assembly of short and long reads. Bioinformatics doi:10.1093/bioinformatics/btv688.
10. Li D, Luo R, Liu CM, Leung CM, Ting HF, Sadakane K, Yamashita H, Lam TW. 2016. MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. Methods 102:3-11.
11. Peng Y, Leung HC, Yiu SM, Chin FY. 2012. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics 28:1420-8.
12. Mikheenko A, Valin G, Prjibelski A, Saveliev V, Gurevich A. 2016. Icarus: visualizer for de novo assembly evaluation. Bioinformatics 32:3321-3323.
13. 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.
14. 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.
15. 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.
16. Meheust R, Burstein D, Castelle CJ, Banfield JF. 2019. The distinction of CPR bacteria from other bacteria based on protein family content. Nat Commun 10:4173.

17. Yakimov MM, Merkel AY, Gaisin VA, Pilhofer M, Messina E, Hallsworth JE, Klyukina AA, Tikhonova EN, Gorlenko VM. 2022. Cultivation of a vampire: 'Candidatus Absconditicoccus praedator'. Environ Microbiol 24:30-49.
18. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res 44:6614-24.
19. Cross KL, Campbell JH, Balachandran M, Campbell AG, Cooper SJ, Griffen A, Heaton M, Joshi S, Klingeman D, Leys E, Yang Z, Parks JM, Podar M. 2019. Targeted isolation and cultivation of uncultivated bacteria by reverse genomics. Nat Biotechnol 37:1314-1321.