Complete Genome Sequence of *Desulfomicrobium* sp. ZS1 from Zodletone Spring in Oklahoma, USA

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

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

Running head: *Desulfomicrobium* sp. ZS1 genome sequence

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

Abstract

Desulfomicrobium sp. ZS1 is an obligate anaerobic, sulfate-reducing member of the Desulfobacterota from Zodletone spring, an anoxic sulfide-rich spring in southwestern Oklahoma. Its complete genome was sequenced using a combination of Illumina and Oxford Nanopore platforms and encodes 3364 proteins and 81 RNAs on a single chromosome.

The anoxic sediments and air exposed water of Zodletone spring (Kiowa County, Oklahoma, GPS coordinates 35.002444 N, 98.688167 W) host some of the most highly diverse microbial communities known, encompassing over 60 phyla and candidate phyla of Bacteria and Archaea (1). The interplay of anaerobic and aerobic sulfur-based metabolism in Zodletone may be reminiscent of physiological adaptations characteristic of microbial life during Archean Earth leading to the great oxygenation event (2). *Desulfomicrobium* sp. strain ZS1 was isolated from a Zodletone sediment sample (1) as a strict anaerobic, motile chemolithoheterotroph, reducing sulfate with lactate as electron donor. A pure culture was obtained following streaking to single colonies on DSMZ medium 63 at 25^{o}C under 85% N_2, 10% CO_2 and 5% H_2.

For genomic DNA isolation, strain ZS1 was grown in 50 ml liquid DSMZ medium 63 for 5 days at 25^{o}C. All subsequent protocols followed manufacturers' instructions. DNA was isolated using the Promega Wizard HMW DNA extraction kit. A short insert library was prepared using the Illumina Nextera XT DNA Library Preparation Kit, followed by sequencing (2x250-nucleotide reads) on a MiSeq instrument (Illumina, Inc., San Diego, CA), yielding 1 million paired reads. Trimmomatic v0.36 (3) was used for quality-based trimming. Long-read sequencing was performed using the Oxford Nanopore Ligation Sequencing Kit followed by sequencing on a MinION R9.4.1 device (Oxford Nanopore Technologies, Inc., Cambridge MA), yielding 1.9 Gb with an N50 of 8.2 kbp. All subsequent data analyses were performed using software defaults. Base calling and long-read polishing were performed using ONT's Guppy and Medaka v1.5,

respectively. First pass assemblies of the long reads were generated using Trycycler v0.5.0 pipeline (4) with Miniasm/Minpolish v0.3-r179 (5), Flye v2.9 (6), Raven v1.5.3 (7) and wtdbg2 v2.5 (8) assemblers, followed by polishing of the consensus single contig with the Illumina short reads using Polypolish v0.4.3 (9) and POLCA v4.0.5 (10) in tandem. Circularity was confirmed as part of Trycycler pipeline assembly and further verified by back mapping of the Illumina reads. The genome is 3,867,579 bp long, with an average coverage of 479x and a G+C% of 58.9.

To predict and annotate the genes, we used the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) v6.1 (11). The genome encodes 3364 proteins and 81 RNAs, which include three ribosomal RNA operons. The gene encoding the chromosomal replication initiator protein DnaA was set as the first gene. Comparative genomic analyses were performed using software implemented in KBase (12) as follows. A phylogenetic tree constructed using SpeciesTree v2.2.0, using a set of 49 core, universal bacterial genes, placed *Desulfomicrobium* sp. ZS1 closest to *Desulfomicrobium baculatum* strain X^T, the genus type species (NCBI accession PRJNA29527)(13, 14)(Fig. 1). Whole-genome average nucleotide identity (ANI) between the two genomes, calculated using FastANI v0.1.3 (15), is of 94%, suggesting that ZS1 is a closely related species to *D. baculatum. Desulfomicrobium* sp. ZS1 will facilitate studies on the evolution of microbial sulfur metabolism and adaptation to anoxic and microoxic environments.

Data availability. The annotated genome sequence has been deposited in GenBank under the accession number CP100351. The version described in this article is CP100351.1. The Nanopore and Illumina reads are available in the NCBI Sequence Read Archive (SRA) under the accession numbers SRR21699889 and SRR20017258, respectively.

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. 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.
2. 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.
3. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114-20.
4. Wick RR, Judd LM, Cerdeira LT, Hawkey J, Meric G, Vezina B, Wyres KL, Holt KE. 2021. Trycycler: consensus long-read assemblies for bacterial genomes. Genome Biol 22:266.
5. Li H. 2016. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. Bioinformatics 32:2103-10.
6. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol 37:540-546.

7. Vaser R, Šikić M. 2021. Time- and memory-efficient genome assembly with Raven. Nature Computational Science 1:332-336.
8. Ruan J, Li H. 2020. Fast and accurate long-read assembly with wtdbg2. Nat Methods 17:155-158.
9. Wick RR, Holt KE. 2022. Polypolish: Short-read polishing of long-read bacterial genome assemblies. PLoS Comput Biol 18:e1009802.
10. Zimin AV, Salzberg SL. 2020. The genome polishing tool POLCA makes fast and accurate corrections in genome assemblies. PLoS Comput Biol 16:e1007981.
11. 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.
12. 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.
13. Anonymous. 1994. Validation of the publication of new names and new combinations previously effectively published outside the IJSB. List No. 49. IntJSystBacteriol 44:370-371.
14. Copeland A, Spring S, Goker M, Schneider S, Lapidus A, Del Rio TG, Tice H, Cheng JF, Chen F, Nolan M, Bruce D, Goodwin L, Pitluck S, Ivanova N, Mavrommatis K, Ovchinnikova G, Pati A, Chen A, Palaniappan K, Land M, Hauser L, Chang YJ, Jeffries CC, Meincke L, Sims D, Brettin T, Detter JC, Han C, Chain P, Bristow J, Eisen JA, Markowitz V, Hugenholtz P, Kyrpides NC, Klenk HP, Lucas S. 2009. Complete genome sequence of Desulfomicrobium baculatum type strain (X). Stand Genomic Sci 1:29-37.
15. Jain C, Rodriguez RL, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun 9:5114.

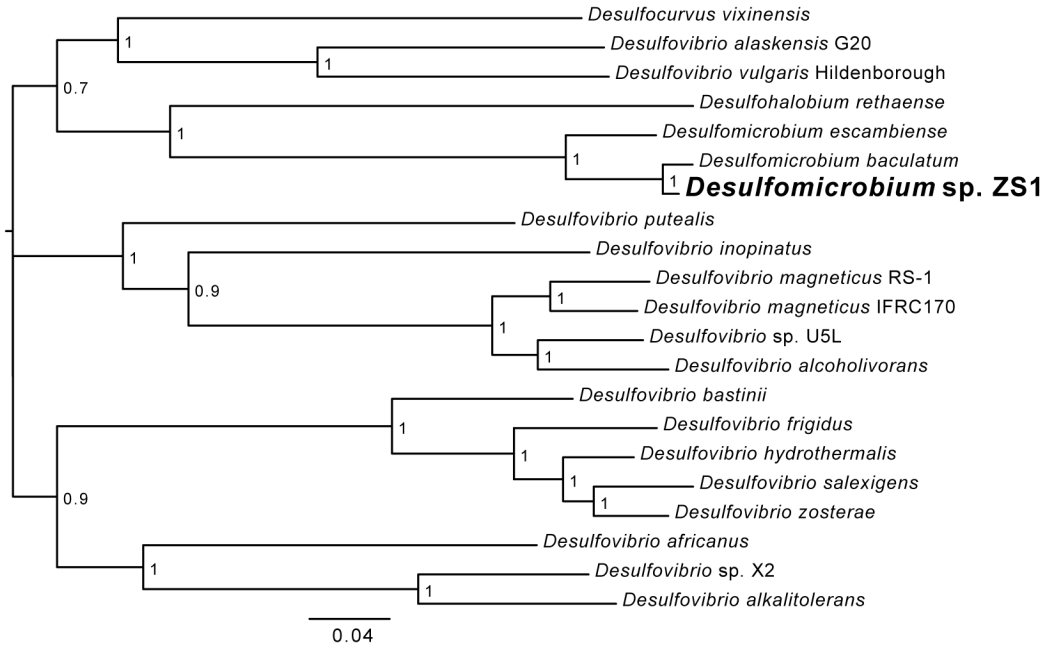

Figure 1. Phylogenetic tree of *Desulfomicrobium* sp. ZS1 and related Desulfobacterota species, based on 49 core, universal bacterial proteins. Numbers at the nodes indicate support values. The scale bar indicates estimated amino acid substitutions per site.