Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Microbial diversity in an early earth analogue: From ancient novelty to modern commonality

C. Ryan Hahn
Okahoma State University
Ibrahim Farag
Okahoma State University
Chelsea Murphy
Okahoma State University
Mircea Podar
Oak Ridge National Laboratory https://orcid.org/0000-0003-2776-0205
Mostafa Elshahed
Oklahoma State University https://orcid.org/0000-0002-1067-1647
Noha Youssef (Noha@Okstate.edu)
Oklahoma State University

Article

Keywords: anaerobic habitats, spring sediments, oxygenation

Posted Date: August 10th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-751738/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Abstract

Life emerged and diversified in the absence of molecular oxygen [1]. The prevailing anoxia and unique sulfur chemistry in the Paleo-, Meso- and Neoarchean, and early Proterozoic eons may have supported microbial communities that are drastically different than those currently thriving on the earth's surface [2–4]. Zodletone spring in southwestern Oklahoma represents a unique habitat where spatial sampling could substitute for geological eons: from the anoxic, surficial light-exposed sediments simulating a preoxygenated earth, to overlaid water column where air exposure simulates the relentless oxygen intrusion during the Neo-Proterozoic [5]. We discovered a remarkably diverse microbial community in the spring sediments, with two thirds (340/516) of the metagenomic assembled genomes belonging to 200 bacterial and archaeal families that were either previously undescribed or are extremely rare elsewhere on earth. Such diversity is underpinned by the widespread occurrence of sulfite-, thiosulfate-, tetrathionate-, and sulfur-reduction, in contrast with a paucity of sulfate-reduction metabolism in those taxa. This greatly expands the diversity of lineages mediating reductive sulfur cycling processes in the tree of life. In the overlaying water community oxygen intrusion leads to the establishment of a significantly less diverse community dominated by well-characterized lineages and the prevalence of oxidative sulfur cycling processes. Such transition from ancient novelty to modern commonality underscores the profound impact of the great oxygenation event on the earth's surficial anoxic community.. It also suggests that novel and rare lineages encountered in current anaerobic habitats could represent taxa once thriving on the anoxic earth that have failed to adapt to the progressive oxygenation.

Main

Sulfur is one of the most abundant elements on earth, exhibiting a wide range of oxidation states (-2 to + 6). Microorganisms have evolved a plethora of genes and pathways for exploiting sulfur-redox reactions for energy generation. Thermodynamic considerations limit reductive sulfur processes to habitats where oxygen is limited. Sulfate is highly abundant on the current earth, and hence sulfate-reduction dominates reductive processes in the global sulfur cycle in permanently and seasonally anoxic and hypoxic habitats in marine [6–8], freshwater [9], terrestrial [10], and subsurface [11] ecosystems. However, during the first two billion years of its history the earth's surface was completely anoxic, and the availability and speciation of various sulfur species greatly differed from current values. Sulfate levels were significantly lower (estimates of < 200 µM-1mM from the Archean up to the Paleoproterozoic; 2.3 Gya) [2,4,5,12], while sulfur-cycle intermediates (SCI) appear to have played an important role in shaping the ancient sulfur cycle [3]. Modeling suggests that mM levels of SO_3^{2-} were attained in the Archean anoxic shallow surficial aquifers as a result of dissolution of the volcanic SO_2 prevailing in aquatic habitats [12]. Isotopic studies have demonstrated the importance of elemental sulfur, sulfite, and thiosulfate reduction in the Archean [3,13].

The evolution of life (3.8 to 4.0 Gya) in the early Archean era and the subsequent evolution of major microbial clades in the Archean and early Proterozoic [1] occurred within this background of anoxia and

sulfur-chemistry. As such, it has been speculated that organisms using intermediate forms of sulfur were likely more common than sulfate-reducing organisms in pre-oxygenated earth [3]. However, while isotopic fractionation, modeling, and microscopic studies provide clues on prevailing sulfur speciation patterns and biological processes, the identity of microorganisms mediating such processes is unknown. This is due to constraints on preservation of nucleic acids and other biological macromolecules, with the oldest successful DNA sequenced sample being only 1.2 M years old [14].

Investigation of modern ecosystems with conditions resembling those prevailing on the ancient earth could provide important clues to the nature, identity, and evolutionary trajectory of microorganisms that once thrived in these eons. In Zodletone spring in southwestern Oklahoma, the prevailing conditions are analogous to those predominant on the earth's surface in the late Archean and early Proterozoic eons (S. text, Figure S1). At the source of the spring, anoxic, surficial, light-exposed conditions are maintained in the sediments by constant emergence of sulfide-saturated water at the spring source from anoxic underground water formations in the Anadarko basin, along with gaseous hydrocarbons, which occur in seeps in the general vicinity. These surficial anoxic conditions also support a sulfur chemistry characterized by high levels of sulfide, sulfite, sulfur (soluble polysulfide), thiosulfate, and a low level of sulfate (S. text). Further, the sediments at the source of the spring are overlaid by an air-exposed water column (60 cm), where oxygen intrusion leads to a vertical oxygen gradient ranging from oxygen saturation in the top 1 µm, to hypoxic in the middle, to anoxic in deeper layers overlaying the sediments.

We combined metagenomic, metatranscriptomic, and amplicon-based approaches to characterize the microbial communities and sulfur cycling processes in Zodletone spring. The anoxic sediments served as a proxy for the late Archean and early Proterozoic communities, while the overlaid hypoxic water pool served as a proxy for oxygen intrusion during the Neoproterozoic. We binned 516 metagenome-assembled genomes (MAGs) from 281 Gbp raw sequence data from the anoxic spring source (Table S1). These genomes belonged to 300 different families in 64 phyla or candidate phyla (53 bacterial and 11 archaeal) (Fig. 1a-b). Diversity assessment utilizing small subunit ribosomal protein S3 from assembled contigs (n = 2079), as well as a complementary 16S rRNA amplicon sequencing (n = 309,074 sequences) identified higher levels of taxonomic diversity, but the overall community composition profiles generated from all three approaches were highly similar (S. text, Figure S2). Assessment of novelty and degree of uniqueness of sediment MAGs identified a remarkably high number of previously undescribed lineages, as well as Lineages exhibiting Rare global Distribution (LRD) patterns compared to other present time earth environments (Fig. 1, 2a, Table S1). For example, at the family level, 132 (26%) and 208 (40%) genomes clustered into 97 novel and 113 LRD families respectively. The bacterial phyla Chloroflexota (n = 69), Planctomycetota (n = 47), Bacteroidota (n = 43), Desulfobacterota (n = 43), Spirochaetota (n = 28 genomes), Patescibacteria (n = 20 genomes), and the archaeal phylum Nanoarchaeota (n = 21) were the most relatively abundant phyla in Zodletone spring sediments, albeit representing only 53% of the total number of recovered genomes (Fig. 1, 2C, S3, S. text). An extreme paucity of genomes belonging to the Proteobacteria (6 genomes) and Firmicutes (12 genomes), the most widely distributed and abundant taxa within current biomes [15], and an absence of the oxygen-generating cyanobacteria was noted (Fig. 1,

Table S1). Previously undescribed and LRD lineages represented the majority of genomes in these most abundant phyla (S. text). For example, within the Chloroflexota, 38/69 genomes belonged to 3 orders and 5 previously undescribed families, as well as multiple LRD orders (Thermoflexales, 4572-78, and UBA2777) and families (E44-bin32, Fen-1058, J111, RBG-13-53-26, RBG-16-64-43, UBA11579, UBA11858, UBA2029, UBA2162, UBA3940, UBA4811, UBA4823, UBA5620, UBA5760, and UBA6092) (S. text, Figure S3).

In contrast, only 114 genomes were binned from 323 Gbp of raw data from the overlaid planktonic water community. Genomes recovered from the water column belonged to a lower number of families (n = 79) and phyla (n = 27) (Table S1). The community exhibited a much lower proportion of previously undescribed and LRD taxa when compared to the sediment community (Fig. 2a-b). Sixty-two genomes belonged to shared families with the sediment community, and 52 were water-specific. Water-specific genomes mostly belonged to well-characterized microbial lineages within the Alphaproteobacteria, Gammaproteobacteria, Camplylobacterota, Firmicutes/Firmicutes_A, Bacteroidota, and Spriochaetota (S. text, Fig. 1). As such, it appears that oxygen intrusion limits the growth of many of the unique microbes prevalent in the sediment, and triggered the proliferation of cosmopolitan lineages within the bacterial tree of life.

Analysis of predicted metabolic capacities and energy conservation pathways in all obtained sediment genomes revealed prevalence of reductive sulfur processes (149/564 genomes) in Zodletone spring sediments communities (S. text, Table S2). Strict fermentative capacities were also widespread, being observed in (100/564 genomes). Strict fermentative lineages mediated the utilization of a wide range of organic substrates, e.g. sugars, amino acids, short chain fatty acids, complex carbohydrates, long chain fatty acids, and short chain alkanes, producing numerous fermentative end products including lactate, formate, acetate, ethanol, succinate, and hydrogen (S. text, Table S2). Primary productivity in the sediments appears to be mostly mediated via hydrogen utilization (69 genomes) coupled to either sulfur-cycle intermediates (SCI) reduction, or to CO_2 fixation by hydrogenotrophic methanogens and acetogens (S. text, Table S2). On the other hand, aerobic, nitrate, or Fe^{3+} respiration, chemolithotrophic nitrification, and photosynthesis potentials were largely absent (Fig. 1, S. text, Table S2).

To investigate sulfur-transformation processes in the spring sediments, we mapped the distribution of key sulfur-cycling genes, and inferred capacities in genomes via the occurrence of entire pathways (Fig. 3). Predicted capacities were subsequently substantiated by examining operon organization and phylogenetic analysis (Fig. 4, S4-S6). Further, time-series metatranscriptomic data were used to identify the metabolically active fraction of the community and document the expression of key pathways (Figure S7). Within the anoxic sediments, a total of 149 genomes (29 % of all genomes), belonging to 32 phyla and 97 families were involved in at least one reductive sulfur processes, while only 21 (4% of all genomes) encoded at least one sulfur oxidation pathway (Fig. 3, Table S3). Sulfate-reduction capacity was observed in only 18 sediment genomes (Figs. 3, 4a), but exhibited a unique community composition, belonging to the Zixibacteria, Acidobacteriota (family UBA6911, equivalent to Acidobacteria group 18), Myxococcota, Bacteroidota (LRD families UBA10428 and UBA5072 within Bacteroidales),

Planctomycetota (1 novel family within class Phycisphaerae), candidate phylum OLB16, as well as LRD and novel lineages within the Desulfobacterota (S. text). Only one genome belonging to the canonical sulfate-reducing family Desulfovibrionaceae was identified. Sulfite (but not sulfate) reduction via the DSR system was identified in only 8 genomes belonging to 7 mostly novel and LRD families within the Planctomycetota, Chloroflexota, and Desulfobacterota (S. text. Figures 3, 4a, S3, Table S3). On the other hand, sulfite-reduction capacity within Zodletone spring sediment solely via the Asr/Hdr system was rampant, being encountered in 104 genomes belonging to 72 (31 novel and 25 LRD) families in 28 phyla (Fig. 3, 4b, S. text, Table S3). For example, within the Chloroflexota, such capacity was encoded in 15 genomes all from uncultured lineages, 6 of which belong to novel families and 6 to LRD families (S. text, Table S3). Further, the capacity was also rampant in a wide range of phyla showing a fairly limited distribution on the current earth, e.g. candidate phyla CSSED10-310, FCPU426, RBG-13-66-14, SM23-31, SZUA-182, UBP14, Aureabacteria, and Sumerlaeota. Interestingly, all genomes belonging to the phylum Krumholzibacteriota, recently described from the spring sediment [16], encoded complete anaerobic sulfite reductase systems. Thiosulfate disproportionation capacity using the *phsABC* system was observed in 11 genomes belonging to 6 phyla, some of which also encoded a dissimilatory sulfite reductase (the Asr system) system for thiosulfate reduction to only hydrogen sulfide, while others encoded the sulfite dehydrogenase SoeABC system for thiosulfate disproportionation to both hydrogen sulfide and sulfate (Figs. 3, S4, S. text, Table S3). Seventy-three Zodletone sediment genomes representing 14 phyla encoded the octaheme tetrathionate reductase (Otr) enzyme, and 68 genomes encoded the Ttr enzyme system for tetrathionate reduction to thiosulfate (Figs. 3, S5, S. text, Table S3). Finally, 20 genomes encoded *psrABC* genes for polysulfide reduction (Fig. 3, S6, S. text, Table S3). Within lineages mediating reductive sulfur processes in Zodletone sediments, a wide range of substrates supporting sulfidogenesis were identified (Table S2, Fig. 3, S. text). Metatranscriptomic analysis identified all S-species reduction/ disproportionation genes discussed above, with transcripts belonging to 51 different phyla identified (S. text, Figure S7).

In contrast, oxidative sulfur processes dominated the water community, while reductive sulfur-processes were extremely sparse (S. text, Fig. 3, Table S3). Pathways encoding sulfide, sulfur, thiosulfate, tetrathionate, and/or sulfite oxidation to sulfate present in 59/114 genomes (52% of all water genomes) belonging to 43 families in 13 phyla were identified. Encoded sulfide and sulfur cycle intermediates (SCI) oxidation pathways included the SOX system (mediating oxidation of a wide range of reduced sulfur-species to sulfate) in the well-characterized families Acidithiobacillaceae, Burkholderiaceae, Halothiobacillaceae, Rhodobacteraceae, and Thiomicrospiraceae (Proteobacteria) and Sulfurimonadaceae (Campylobacterota), the sulfide dehydrogenase fccAB and/or the sulfide:quinone oxidoreductase Sqr systems for sulfide oxidation to sulfur/polysulfide identified mostly in the Chlorobiaceae, Prolixibacteraceae, and Paludibacteracaeae (Bacteroidota), Acidithiobacillaceae, Burkholderiaceae, Chromatiaceae, Halothiobacillaceae, Methylothermaceae, Rhodobacteraceae, Thiomicrospiraceae (Proteobacteria), and Sulfurimonadaceae, Sulfurospirillaceae, Sulfurovaceae (Campylobacterota), and sulfite oxidation to sulfate via the reversal of AprAB + QmoABC system, the sulfite dehydrogenase (quinone) SoeABC, or the sulfite dehydrogenase (cytochrome) SorAB identified in

26 genomes of mostly well characterized families within the Bacteroidales, Proteobacteria, and Campylobacterota. Finally, eight water genomes from well-characterized lineages encoded thiosulfate to tetrathionate oxidation capacities via either the thiosulfate dehydrogenase *tsdA* [EC: 1.8.2.2] or the thiosulfate dehydrogenase (quinone) *doxAD* [EC: 1.8.5.2] (S. text, Table S3).

Collectively, phylogenomic analysis documents a unique microbial community in the spring anoxic sediments dominated by previously unencountered or extremely rare taxa on the current earth. Metabolic analysis suggests that such unique community is sustained by respiration and disproportionation of a wide range of SCI abundant in the spring. We posit that the high level of diversity and the abundance of previously undescribed and LRD lineages within Zodletone spring sediment S-reducing microbial (SRM) community could be attributed to two factors. First, the availability of a wide range of sulfur cycle intermediates selects for a more diverse community of SRM in the spring, when compared to predominantly sulfate-driven marine and freshwater ecosystems. Second, additional factors constraining SRM growth in several habitats, e.g. diel or seasonal intrusion of oxygen, Fe and NO_3 [6,17–19], recalcitrance of available substrates [11,20,21], temperature [22,23], pH [24–26], salinity [27], and pressure extremes [20,28], or combinations thereof, are absent in the spring.

The prevailing conditions in Zodletone spring (anaerobic, surficial, light-exposed, sulfidic, with abundance of SCI) remarkably mimic the conditions that prevailed in the late Archean and early Proterozoic eons. Due to the sensitivity and expected lack of adaptive mechanisms to cope with atmospheric oxygen in multiple strict anaerobes, as well as the chemical instability of multiple S species in an oxygenated atmosphere, the Great Oxidation Event (GOE) exerted a profound negative impact on anaerobic surficial life forms (the oxygen catastrophe) leading to the first and arguably most profound extinction event in earth's history. This study infers that the microbial communities presumably thriving in anoxic surficial earth were extremely diverse with an abundance of SRM lineages. We argue that many such lineages were driven to extreme rarity in current environments, a reflection of their failure to adapt to the rise of atmospheric oxygen and the subsequent associated changes in earth's redox status and sulfur chemistry.

In addition to suppressing anaerobiosis in atmospherically-exposed habitats, the GOE also led to a significant change in the S cycle, from one based on atmospheric inputs (volcanic SO_2 release and dissolution in aqueous habitats) to one dependent on oxidative weathering. Such transition has led to the release of large amounts of sulfate derived from the oxidation of pyrite and the dissolution of sulfate minerals [29], hitherto a minor byproduct of Archean abiotic and biotic reactions [3,30]. Our analysis comparing anoxic sediment communities to overlaid water hypoxic communities suggest that oxygen intrusion and loss of niches associated with geological transformations have triggered a mass extinction/confinement of a wide swath of branches within the bacterial tree, and greatly altered the microbial community mediating reductive sulfur transformation processes on earth.

In summary, by examining microbial diversity in Zodletone spring, we greatly expand the overall diversity within the tree of life via the discovery and characterization of a numerous previously uncharacterized lineages; and significantly enrich representation of a multitude of LRD lineages. We also describe a

unique sulfur-cycling community in the spring that is largely dependent on sulfite, thiosulfate, sulfur, and tetrathionate, rather than sulfate, as an electron acceptor. Given the inferred similarity to conditions prevailing prior to the GOE, we consider the spring an invaluable portal to investigate microbial communities and processes thriving on the earth's surface during these eons. Furthermore, we posit that the GOE precipitated the near extinction of a wide range of phylogenetically distinct oxygen-sensitive lineages and drastically altered the microbial reductive sulfur-cycling community from sulfite, sulfur, and thiosulfate reducers to predominantly sulfate reduction.

Declarations

Acknowledgments. This work was supported by the National Science Foundation Grants 2016423 (to NHY and MSE) and 2016371 to MP.

Authors' contributions: CRH: Formal analysis, investigation, and visualization; IFF: Formal analysis; CLM: Investigation, MP: Formal analysis, writing review and editing, project administration, and funding acquisition. MSE: Conceptualization, formal analysis, resources, writing original draft, supervision, project administration, funding acquisition. NHY: Conceptualization, formal analysis, resources, writing review and editing, visualization, project administration, and funding acquisition.

Competing interests: The authors declare no competing interest.

Data availability. The whole genome shotgun project was submitted to GenBank under Bioproject ID PRJNA690107 and Biosample IDs SAMN17269717 (for the sediment metagenome) and SAMN17269718 (for the water metagenome). The individual assembled MAGs have been deposited at DDBJ/ENA/GenBank under the accession JAFFZZ000000000-JAFGPI000000000. The version described in this paper is version JAFFZZ010000000-JAFGPI010000000. Metagenomic raw reads for the sediment, and the water are available under SRA accession SRX9813571, and SRX9813572. RNA-seq reads generated in this study are available under SRA accessions SRX9810743, SRX9810744, and SRX9810745 for the morning, afternoon, and evening samples.

Supplementary Information is available for this paper.

References

1 Marin, J. *et al.* The timetree of Prokaryotes: New insights into their evolution and speciation. *Mol. Biol. Evol.* **34**, 437-446 (2016).

2 Canfield, D. E. & Raiswell, R. The evolution of the sulfur-cycle. *Am. J. Sci.* **299**, 697-723 (1999).

3 Habicht, K. S., Canfield, D. E. & Rethmeier, J. Sulfur isotope fractionation during bacterial reduction and disproportionation of thiosulfate and sulfite. *Geochimica et Cosmochimica Acta* **62**, 2585-2595 (1998).

4 Habicht, K. S., M. Gade, Thamdrup, B., Berg, P. & Canfield, D. E. Calibration of sulfate levels in the Archean ocean. *Science* **298**, 2372-2374 (2002).

5 Canfield, D. E., Habicht, K. S. & Thamdrup, B. The Archean Sulfur Cycle and the Early History of Atmospheric Oxygen. *Science* **299**, 658-661 (2000).

6 Jørgensen, B. B., Findlay, A. J. & Pellerin, A. The biogeochemical sulfur-cycle in marine sediments. *Front. Microbiol.* **10**, 849 (2019).

7 Wasmund, K., Mußmann, M. & Loy, A. The life sulfuric: microbial ecology of sulfur cycling in marine sediments. *Environ. Microbiol. Rep.* **9**, 323-344 (2017).

8 van Vliet, D. M. *et al.* The bacterial sulfur cycle in expanding dysoxic and euxinic marine waters. *Environ. Microbiol.* **In Press** (2020).

9 Holmer, M. & Storkholm, P. Sulphate reduction and sulphur cycling in lake sediments: a review. *Freshw. Biol.* **46**, 431-451 (2001).

10 Schmalenberger, A., Drake, H. L. & Küsel, K. High unique diversity of sulfate-reducing prokaryotes characterized in a depth gradient in an acidic fen. *Environ Microbiol* **9**, 1317–1328 (2007).

11 Gieg, L. M., Davidova, I. A., Duncan, K. E. & Suflita, J. M. Methanogenesis, sulfate reduction and crude oil biodegradation in hot Alaskan oilfields. *Environ Microbiol* **12**, 3074-3086 (2010).

12 Ranjan, S., Todd, Z. R., Sutherland, J. D. & Sasselov, D. D. Sulfidic anion concentrations on early earth for surficial origins-of-Life chemistry. *Astrobiology* **18**, 1023-1041 (2018).

13 Philippot, P. *et al.* Early archaean microorganisms preferred elemental sulfur, not sulfate. *Science* **317**, 1534-1537 (2007).

14 van der Valk, T. *et al.* Million-year-old DNA sheds light on the genomic history of mammoths. *Nature* **In Press, https://doi.org/10.1038/s41586-021-03224-9** (2021).

15 Nayfach, S. *et al.* A genomic catalog of Earth's microbiomes. *Nat Biotechnol*, doi:10.1038/s41587-020-0718-6 (2020).

16 Youssef, N. H. *et al.* 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. *Systematic and Applied Microbiology* **42**, 85-93, doi:https://doi.org/10.1016/j.syapm.2018.11.002 (2019).

17 Achtnich, C., Bak, F. & Conrad, R. Competition for electron donors among nitrate reducers, ferric iron reducers, sulfate reducers, and methanogens in anoxic paddy soil. *Biol Fertil Soils* **19**, 65-72 (1995).

18 Glombitza, C., Egger, M., Røy, H. & Jørgensen, B. B. Controls on volatile fatty acid concentrations in marine sediments (Baltic Sea). *Geochimica et Cosmochimica Acta*, 226-241 (2019).

19 Morrison, J. M. *et al.* Spatiotemporal analysis of microbial community dynamics during seasonal stratification events in a freshwater lake (Grand Lake, OK, USA). *PLOS ONE* **12**, e0177488 (2017).

20 Daly, R. A. *et al.* Microbial metabolisms in a 2.5-km-deep ecosystem created by hydraulic fracturing in shales. *Nature Microbiol.* **1**, 16146 (2016).

21 Jørgensen, B. B. & Marshall, I. P. G. Slow microbial life in tthe seabed. *Ann. Rev. Mar. Sci.* **8**, 311-332 (2016).

22 Jørgensen, B. B., Isaksen, M. F. & Jannasch, H. W. Bacterial sulfate reduction above 100°C in deep-sea hydrothermal vent sediments. *Science* **1756-1757** (1992).

23 Scholze, C., Jørgensen, B. B. & Røy, H. Psychrophilic properties of sulfate-reducing bacteria in Arctic marine sediments. *Limnol. Oceanog.* **In Press** https://doi.org/10.1002/lno.11586 (2020).

24 Sánchez-Andrea, I., Sanz, J. L., Bijmans, M. F. & Stams, A. J. Sulfate reduction at low pH to remediate acid mine drainage. *J Hazard Mater* **269**, 98-109, doi:10.1016/j.jhazmat.2013.12.032 (2014).

25 Sorokin, D. Y., Kuenen, J. G. & Muyzer, G. The microbial sulfur cycle at extremely haloalkaline conditions of soda lakes. *Front. Microbiol.* **2**, 44 (2011).

26 Vavourakis, C. D. *et al.* Metagenomes and metatranscriptomes shed new light on the microbial-mediated sulfur cycle in a Siberian soda lake. *BMC Microbiol.* **17**, 69 (2019).

27 Teske, A. *et al.* Sulfate-Reducing Bacteria and Their Activities in Cyanobacterial mats of Solar Lake (Sinai, Egypt). *Appl. Environ. Microbiol.* **1998**, 2943-2951 (1998).

28 Bell, E. *et al.* Active sulfur cycling in the terrestrial deep subsurface. *The ISME J.* **14**, 1260-1272 (2020).

29 Killingsworth, B. A. *et al.* Constraining the rise of oxygen with oxygen isotopes. *Nature Comm.* **10**, 4924 (2019).

30 Holland, H. D. Some applications of thermochemical data to problems of ore deposits; part 2, mineral assemblages and the com- position of ore forming fluids. *Econ. Geol.* **60**, 1101–1166. (1995).

31 Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLOS ONE* **5**, e9490, doi:10.1371/journal.pone.0009490 (2010).

32 Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* **36**, 1925-1927, doi:10.1093/bioinformatics/btz848 (2019).

33 Nakamura, T., Yamada, K. D., Tomii, K. & Katoh, K. Parallelization of MAFFT for large-scale multiple sequence alignments. *Bioinformatics* **34**, 2490-2492, doi:10.1093/bioinformatics/bty121 (2018).

34 Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313, doi:10.1093/bioinformatics/btu033 (2014).

35 Guy, L., Kultima, J. R. & Andersson, S. G. E. genoPlotR: comparative gene and genome visualization in R. *Bioinformatics (Oxford, England)* **26**, 2334-2335, doi:10.1093/bioinformatics/btq413 (2010).

Methods

Site description and geochemistry. Zodletone spring is located in the Anadarko Basin of western Oklahoma (N34.99562° W98.68895°). The spring arises from underground, where water is pumped out slowly along with sediments. Sediments settled at the source of the spring, a boxed square $1m^2$ (Figure S1) are overlaid with water that collects and settles in a concrete pool erected in the early 1900s. The settled water is 50-cm deep above the sediments and is exposed to atmospheric air. Water and sediments originating from the spring source are highly reduced due to the high dissolved sulfide levels (8-10 mM) in the spring sediments. Microsensor measurements show a completely anoxic (oxygen levels < 0.1 µM) and highly reduced source sediments. Oxygen levels slowly increase in the overlaid water column from 2–4 µM at the 2 mm above the source to complete oxygen exposure on the top of the water column [36]. The spring geochemistry has regularly been monitored during the last two decades [36-38] and is remarkably stable. The spring is characterized by low levels of sulfate (50-94 µM), with higher levels of sulfite (0.21 mM), elemental sulfur (0.1 mM), and thiosulfate (0.52) [38,39].

Sampling. Samples were collected from the source sediments and standing overlaid water in sterile containers and kept on ice until brought back to the lab (~2h drive), where they were immediately processed. For metatranscriptomics, samples were collected at three different time points: morning (9:15 am), afternoon (2:30 pm), and evening (5:30 pm) in June 2019; stored on dry ice until transferred to the lab where they were stored at -80ºC until processed for RNA extraction within a week.

Nucleic acid extraction. DNA was directly extracted from 0.5 grams of source sediments. For water samples, water was filtered on 0.2 µm sterile filters. DNA was directly extracted from filters (20 filters, 10 L of water samples). Extraction was conducted using the DNeasy PowerSoil kit (Qiagen, Valencia, CA, USA). RNA was extracted from 0.5 g sediment samples using RNeasy PowerSoil Total RNA Kit (Qiagen, Valencia, CA, USA) according to the manufacturer's instructions.

16S rRNA gene amplification, sequencing, and analysis. Triplicate DNA extractions were performed for both sediment and water samples from the Zoddletone spring. To characterize the microbial diversity based on 16S rRNA gene sequences we used the Quick-16S™ NGS Library Prep Kit (Zymo Research, Irvine CA), following the manufacturer's protocol. For amplification of the V4 hypervariable region we used a mix of modified versions of primers 515F-806R [40], tailored to provide better coverage for several under-

represented microbial lineages. They included 515FY (5'GTGYCAGCMGCCGCGGTAA) [41], 515F-Cren (5' GTGKCAGCMGCCGCGGTAA, for Crenarchaeota) [42], 515F-Nano (5'GTGGCAGYCGCCRCGGKAA, for Nanoarchaeota) [42], 515F-TM7 (5' GTGCCAGCMGCCGCGGTCA for TM7/Saccharibacteria) [43] as forward mix and 805RB (5' GGACTACNVGGGTWTCTAAT) [44] and 805R-Nano (5'GGAMTACHGGGGTCTCTAAT, for Nanoarchaeota) [42] as reverse mix. Purified barcoded amplicon libraries were sequenced on an Illumina MiSeq instrument (Illumina Inc., San Diego, CA) using a v2 500 cycle kit, according to manufacturer's protocol. Demultiplexed forward and reverse reads were imported as paired fastq files into QIIME2 v. 2020.8 [45] for analysis. The DADA2 plugin was used to trim, denoise, pair, purge chimeras and select amplicon sequence variants (ASVs), using the command "qiime dada2 denoise-paired". Between 44k and 194k non-chimeric sequences were obtained for the individual samples. The ASVs were taxonomically classified in QIIME2 using a trained classifier built based on Silva-138-99 rRNA sequence database. The ASVs were assigned to 1643 taxonomic categories corresponding to taxonomic level 7 (species and above) and to 932 genera (level 6). There were no dominating species or genera in either the water or sediment: in the water sample only three taxa reached 3-5% relative abundance, while in the sediment, only three taxa accounted for 2-4% of the community, with 80% of the species being less that 0.1% of the community. Alpha rarefaction curves indicated saturation of observed sequence features (ASVs) at a sequencing depth of 70-80k sequences, the combined number of sequences being 514510 for water and 309383 for sediment.

Metagenome sequencing, assembly, and binning. Metagenomic sequencing was conducted using the services of a commercial provider (Novogene, Beijing, China) using two lanes of the Illumina HiSeq 2500 system for each of the water and sediment samples. Transcriptomic sequencing using Illumina HiSeq 2500 2 × 150bp paired-end technology was conducted using the services of a commercial provider (Novogene Corporation, Beijing, China). Metagenomic reads were assessed for quality using FastQC followed by quality filtering and trimming using Trimmomatic v0.38 [46]. High quality reads were assembled into contigs using MegaHit (v.1.1.3) with minimum Kmer of 27, maximum kmer of 127, Kmer step of 10, and minimum contig length of 1000 bp. Bowtie2 was used to calculate sequencing coverage of each contig by mapping the raw reads back to the contigs. Assembled contigs were searched for ribosomal protein S3 (rpS3) sequences using a custom hidden Markov model (HMM) built from Uniprot reference sequences assigned to the Kegg Orthologies K02982, and K02984 (corresponding to the bacterial, and archaeal RPS3, respectively) using hmmbuild (HMMER 3.1b2). rpS3 Sequences were clustered at 99% ID using CD-HIT as previously suggested for a putative species cutoff for rpS3 data [47]. Taxonomic affiliations of (rpS3) groups were identified using Diamond Blast against the GTDB r95 database [48].

Contigs from the sediment and water assemblies were binned into draft genomes using both Metabat [49] and MaxBin2 [50]. DasTool was used to select the highest quality bins from each metagenome assembly [51]. CheckM was used for estimation of genome completeness, strain heterogeneity, and contamination [52]. Genomic bins showing contamination levels higher than 10%, were further refined based on the

taxonomic affiliations of the binned contigs, as well as the GC content, tetranucleotide frequency, and coverage levels using RefineM [53]. Low quality bins (>10% contamination) were cleaned by removal of the identified outlier contigs, and the percentage completeness and contamination were again re-checked using CheckM.

Genomes classification, annotation, and metabolic analysis. Taxonomic classifications followed the Genome Taxonomy Database (GTDB) release r95 [48], and were carried out using the classify_workflow in GTDB-Tk (v1.1.0) [32]. Phylogenomic analysis utilized the concatenated alignment of a set of 120 single-copy bacterial genes, and 122 single-copy archaeal genes [48] generated by the GTDB-Tk. Maximum-likelihood phylogenomic tree was constructed in FastTree using the default parameters [31].

Annotation and metabolic analysis. Protein-coding genes in genomic bins were predicted using Prodigal [54]. GhostKOALA [55] was used for the functional annotation of every predicted open reading frame in every genomic bin and to assign protein-coding genes to KEGG orthologies (KOs).

Analysis of sulfur cycling genes. To identify taxa mediating key sulfur-transformation processes in the spring sediments, we mapped the distribution of key sulfur-cycling genes in all genomes and deduced capacities in individual genomes by documenting the occurrence of entire pathways (as explained below in details). This was subsequently confirmed by phylogenetic analysis and examining contiguous genes organization in processes requiring multi-subunit and/or multi-gene. Further, expression data was used from three time points to identify the fraction of the community that is metabolically actively involved in the process. Analysis of S cycling capabilities was conducted on individual genomic bins by building and scanning hidden markov model (HMM) profiles as explained below. To build the sulfur-genes HMM profiles, Uniprot reference sequences for all genes with an assigned KO number were downloaded, aligned using Clustal-omega [56], and the alignment was used to build an HMM profile using hmmbuild (HMMER 3.1b2) [57]. For genes not assigned a KO number (e.g. *otr*, *tsdA*, *tetH*), a representative protein was compared against the KEGG Genes database using Blastp and significant hits (those with e-values < e-80) were downloaded and used to build HMM profiles as explained above. The custom-built HMM profiles were then used to scan the analyzed genomes for significant hits using hmmscan (HMMER 3.1b2) [57] with the option -T 100 to limit the results to only those profiles with an alignment score of at least 100. Further confirmation was achieved through phylogenetic assessment and tree building procedures, in which potential candidates identified by hmmscan were aligned to the reference sequences used to build the custom HMM profiles using Clustal-omega [56], followed by maximum likelihood phylogenetic tree construction using FastTree [31]. Only candidates clustering with reference sequences were deemed true hits and were assigned to the corresponding KO.

Sulfate-reduction. Sulfate reduction capacity was assessed by the presence of genes encoding the enzymes 3'-phosphoadenosine 5'-phosphosulfate synthase [Sat; EC:2.7.7.4 2.7.1.25] for sulfate activation to adenylyl sulfate (APS), the enzyme complex adenylylsulfate reductase [AprAB; EC:1.8.99.2] for APS reduction to sulfite, the quinone-interacting membrane-bound oxidoreductase complex [QmoABC]

for electron transfer, the enzyme dissimilatory sulfite reductase [DsrAB; EC:1.8.99.5] and its co-substrate DsrC for dissimilatory sulfite reduction to sulfide, and the sulfite reduction-associated membrane complex DsrMKJOP for linking cytoplasmic sulfite reduction to energy conservation.

Sulfite-reduction. Sulfite could be utilized by most sulfate-reducing microorganisms [58]. Dedicated sulfite-reduction capacity was assessed by the presence of the dissimilatory sulfite reductase system explained above [59,60] with the lack of sulfate-activation (Sat) and reduction (Apr) genes. In addition, sulfite-reduction was assessed via the sole or co-occurrence of the anaerobic sulfite reductase (AsrABC) system [61], along with the membrane-bound associated complex (HdrABC) for transfer of electrons to the AsrC subunit [62]. The Asr enzyme has been shown to function in the cytoplasm in *Salmonella typhimurium* to reduce the sulfite released from respiratory reduction of tetrathionate and thiosulfate [61]. However, a scenario where the Asr enzyme is involved in sulfite respiration is possible via electron transfer from a membrane-bound associated complex to AsrC (the physiological partner of AsrAB). A plausible candidate for this membrane complex is the heterodisulfide reductase-related enzymes (HdrABC), analogous to what was suggested for DsrC (the physiological partner of DsrAB) in organisms lacking the sulfite reduction-associated membrane complex DsrMKJOP [62].

Polysulfide reduction: In addition to sulfate and sulfite, Zodletone spring is euxinic with extremely high levels of zero valent sulfur, available as soluble polysulfide. Respiratory polysulfide reduction was assessed via the identification of the membrane-bound molybdoenzyme complex PsrABC, which reduces polysulfides with electrons obtained from either a hydrogenase or a formate dehydrogenase through a quinone electron carrier [63]. In addition to the membrane-bound Psr system, representatives of the cytolpasmic sulfurhydrogenase I (HydABCD system), and/or II (ShyABCD system) were identified. However, although these enzymes have been shown to be dissimilatory in the archaeon *Pyrococcus furiosus* [64,65], their involvement in an ETS-associated respiration is currently unclear.

Thiosulfate reduction/ disproportionation: Thiosulfate occurs in natural environments as a result of the reaction of sulfite with bisulfide (HS^-) [66]. Thiosulfate is relatively stable at neutral pH and is present in high levels in Zodletone spring, Thiosulfate contains two sulfur atoms: a sulfone-sulfur (oxidation state +5), and a sulfane-sulfur (oxidation state -1). As such, thiosulfate can be disproportionated where the sulfone-sulfur is reduced (serves as an electron acceptor), and the sulfane-sulfur is oxidized (serves as an electron donor), with the products being hydrogen sulfide, and sulfite, respectively. We searched for genes encoding the three known pathways for thiosulfate-disproportionation. First, in pure cultures of several sulfate reducers in the Desulfobacterota and Firmicutes, e.g. *Desulfovibrio*, *Desulfotomaculum*, thiosulfate disproportionation is known to occur via a cytochrome c-dependent thiosulfate reductase [EC: 1.8.2.5] [67-74]. Second, in pure culture members of the family Enterobacteriaceae (Gammaproteobacteria), thiosulfate disproportionation is known to occur via the quinone-dependent membrane-bound molybdopterin-containing thiosulfate reductase PhsABC [75]. Finally, thiosulfate disproportionation to sulfite and hydrogen sulfide can also occur via a rhodanase-like enzyme [EC: 2.8.1.1 or EC: 2.8.1.3], as

shown for several bacterial lineages [76-80], although this could be part of a thiosulfate assimilatory pathway as recently shown in *E. coli* [81].

Following the disproportionation of thiosulfate to sulfite and hydrogen sulfide, microorganisms differ in the fate of the produced sulfite. Some microorganisms reduce the released sulfite to sulfide via a Dsr or Asr dissimilatory sulfite reductase [75]), leading to complete reduction of one thiosulfate molecule to two sulfides (thiosulfate-reduction). Others oxidize the released sulfite to sulfate via the reversal of the sulfate reduction pathway [74,82], or via the sulfite dehydrogenases SorAB or SoeABC [83], leading to the final conversion of one thiosulfate molecule to one sulfide and one sulfate molecules. The distribution of all thiosulfate disproportionation capacities were assessed by the occurrence of one of the three pathways described above, and the fate of sulfite in genomes mediating the initial disproportionation steps was assessed as described above.

Tetrathionate reduction: Tetrathionate has two sulfur atoms in oxidation state of 0 while the other two are in oxidation state of +5. In nature, tetrathionate is formed via the biotic or abiotic oxidation of thiosulfate under anoxic conditions [66]. Some microorganisms are capable of tetrathionate respiration via membrane-bound tetrathionate reductases that will reduce tetrathionate to thiosulfate serving as the terminal oxidase in a short electron transport system. Enzymes mediating such process include octaheme tetrathionate reductase (otr) [84], as well as the guanylyl molybdenum cofactor-containing tetrathionate reductase (ttrABC) [85]. The produced thiosulfate could be metabolized through disproportionation as described above.

Oxidative sulfur processes. The versatile sulfur oxidation (SOX enzyme complex) system was assessed in all genomes. The SOX system mediates the oxidation of a wide range of reduced sulfur compounds (sulfide, sulfite, thiosulfate, and elemental sulfur) directly to sulfate. Sulfide oxidation to sulfur was also assessed by the presence of the sulfide dehydrogenase FccAB [EC: 1.8.2.3] and/or the sulfide:quinone oxidoreductase Sqr [EC: 1.8.5.4], both known to oxidize sulfide to sulfur or polysulfide. Sulfur/polysulfide oxidation to sulfite was assessed via the reversal of the Dsr system (encompassing the full Dsr system *dsrAB+dsrC+dsrMKJOP*, in addition to the genes *dsrEFH*, *tusA*, and *rhdA*). Sulfite oxidation to sulfate was assessed via the reversal of AprAB+QmoABC system, the sulfite dehydrogenase (quinone) SoeABC [EC: 1.8.5.6], or the sulfite dehydrogenase (c-type cytochrome) SorAB [EC: 1.8.2.1]. Thiosulfate oxidation to tetrathionate was assessed via the thiosulfate dehydrogenase *tsdA* [EC: 1.8.2.2], or the thiosulfate dehydrogenase (quinone) *doxAD* [EC: 1.8.5.2]. Tetrathionate generated could be cleaved using tetrathionate hydrolase (*tetH*) [86] that is known to cleave tetrathionate to thiosulfate, sulfur, and sulfate, or converted to sulfite using the rDSR system.

Phylogenetic analysis and operon organization of S cycling genes. The phylogenetic affiliation of the S cycling proteins AsrB, Otr, PhsC, PsrC, and DsrAB was examined by aligning Zodletone genome predicted protein sequences to Uniprot reference sequences using Mafft [87]. The DsrA and DsrB alignments were concatenated in MEGA X [88]. All alignments were used to construct maximum likelihood phylogenetic

trees in RAxML [89]. The R package genoPlotR [35] was used to produce gene maps for the DSR and ASR loci in Zodletone genomes using the Prodigal predicted gene starts, ends, and strand.

Transcription of sulfur cycling genes. A total of 21.4 M, 27.9 M, and 22.5 M 150-bp paired-end reads were obtained from the morning, afternoon, and evening RNA-seq libraries. Reads were pseudo-aligned to all Prodigal-predicted genes from all genomes using Kallisto with default settings [90]. The calculated transcripts per million (TPM) were used to obtain total transcription levels for genes identified from genomic analysis as involved in S cycling in the spring.

Additional metabolic analysis. For all other non-sulfur related functional predictions, combined GhostKOALA outputs of all genomes belonging to a certain order (for orders with 5 genomes or less; n=206), or family (for orders with more than 5 genomes; n=85) were checked for the presence of groups of KOs constituting metabolic pathways (additional file 1). The list of these 291 lineages is shown in Table S2. The presence of at least 80% of KOs assigned to a certain pathway in at least one genome belonging to a certain order/family was used as an indication of the presence of that pathway in that order/family. Such criteria were used for the prediction of autotrophic capabilities, as well as catabolic heterotrophic degradation capabilities of sugars, amino acids, long-chain fatty acids, short chain fatty acids, anaerobic benzoate degradation, anaerobic short chain alkane degradation, aerobic respiration, nitrate reduction, nitrification, and chlorophyll biosynthesis. Glycolytic, and fermentation capabilities were predicted by feeding the GhostKOALA output to KeggDecoder [91]. Proteases, peptidases, and protease inhibitors were identified using Blastp against the Merops database [92], while CAZymes (glycoside hydrolases [GHs], polysaccharide lyases [PLs], and carbohydrate esterases [CEs]) were identified by searching all ORFs from all genomes against the dbCAN hidden Markov models V9 [93] (downloaded from the dbCAN web server in September 2020) using hmmscan. FeGenie [94] was used to predict the presence of iron reduction and iron oxidation genes in individual bins.

References

36 Buhring, S. I. *et al.* 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**, 166-179 (2011).

37 Elshahed, M. S. *et al.* Bacterial diversity and sulfur cycling in a mesophilic sulfide-rich spring. *Appl. Environ. Microbiol.* **69**, 5609-5621 (2003).

38 Senko, J. M. *et al.* Barite deposition mediated by photootrophic sulfide-oxidizing bacteria. *Geochim. Cosmochim. Acta* **68**, 773-780 (2004).

39 Spain, A. M., Najar, F. Z., Krumholz, L. R. & Elshahed, M. S. Metatranscriptomic analysis of a high-sulfide aquatic spring reveals insights into sulfur cycling and unexpected aerobic metabolism. *Peer J* **3**,

e1259 (2015).

40 Caporaso, J. G. *et al.* Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences* **108**, 4516-4522, doi:10.1073/pnas.1000080107 (2011).

41 Parada, A. E., Needham, D. M. & Fuhrman, J. A. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environmental Microbiology* **18**, 1403-1414, doi:https://doi.org/10.1111/1462-2920.13023 (2016).

42 Podar, P. T., Yang, Z., Björnsdóttir, S. H. & Podar, M. Comparative Analysis of Microbial Diversity Across Temperature Gradients in Hot Springs From Yellowstone and Iceland. *Frontiers in Microbiology* **11** (2020).

43 Cross, K. L. *et al.* Targeted isolation and cultivation of uncultivated bacteria by reverse genomics. *Nature Biotechnology* **37**, 1314-1321, doi:10.1038/s41587-019-0260-6 (2019).

44 Walters, W. *et al.* Improved Bacterial 16S rRNA Gene (V4 and V4-5) and Fungal Internal Transcribed Spacer Marker Gene Primers for Microbial Community Surveys. *mSystems* **1**, e00009-00015, doi:10.1128/mSystems.00009-15 (2016).

45 Bolyen, E. *et al.* Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology* **37**, 852-857, doi:10.1038/s41587-019-0209-9 (2019).

46 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)* **30**, 2114-2120, doi:10.1093/bioinformatics/btu170 (2014).

47 Diamond, S. *et al.* Mediterranean grassland soil C–N compound turnover is dependent on rainfall and depth, and is mediated by genomically divergent microorganisms. *Nature Microbiology* **4**, 1356-1367, doi:10.1038/s41564-019-0449-y (2019).

48 Parks, D. H. *et al.* A complete domain-to-species taxonomy for Bacteria and Archaea. *Nature Biotechnology* **38**, 1079-1086, doi:10.1038/s41587-020-0501-8 (2020).

49 Kang, D. D. *et al.* MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359-e7359, doi:10.7717/peerj.7359 (2019).

50 Wu, Y.-W., Simmons, B. A. & Singer, S. W. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* **32**, 605-607, doi:10.1093/bioinformatics/btv638 (2015).

51 Sieber, C. M. K. *et al.* Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nature Microbiology* **3**, 836-843, doi:10.1038/s41564-018-0171-1 (2018).

52 Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* **25**, 1043-1055, doi:10.1101/gr.186072.114 (2015).

53 Parks, D. H. *et al.* Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nature Microbiology* **2**, 1533-1542, doi:10.1038/s41564-017-0012-7 (2017).

54 Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119-119, doi:10.1186/1471-2105-11-119 (2010).

55 Kanehisa, M., Sato, Y. & Morishima, K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *Journal of Molecular Biology* **428**, 726-731, doi:https://doi.org/10.1016/j.jmb.2015.11.006 (2016).

56 Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539-539, doi:10.1038/msb.2011.75 (2011).

57 Mistry, J., Finn, R. D., Eddy, S. R., Bateman, A. & Punta, M. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic acids research* **41**, e121, doi:10.1093/nar/gkt263 (2013).

58 Rabus, R., Hansen, T. A. & Widdel, F. in *The Prokaryotes: Prokaryotic Physiology and Biochemistry* (eds Eugene Rosenberg *et al.*) 309-404 (Springer Berlin Heidelberg, 2013).

59 Hausmann, B. *et al.* Peatland Acidobacteria with a dissimilatory sulfur metabolism. *ISME J* **12**, 1729-1742, doi:10.1038/s41396-018-0077-1 (2018).

60 Colman, D. R. *et al.* Phylogenomic analysis of novel Diaforarchaea is consistent with sulfite but not sulfate reduction in volcanic environments on early Earth. *ISME J* **14**, 1316-1331, doi:10.1038/s41396-020-0611-9 (2020).

61 Huang, C. J. & Barrett, E. L. Sequence analysis and expression of the *Salmonella typhimurium* asr operon encoding production of hydrogen sulfide from sulfite. *J Bacteriol* **173**, 1544-1553, doi:10.1128/jb.173.4.1544-1553.1991 (1991).

62 Venceslau, S. S., Stockdreher, Y., Dahl, C. & Pereira, I. A. The "bacterial heterodisulfide" DsrC is a key protein in dissimilatory sulfur metabolism. *Biochim Biophys Acta* **1837**, 1148-1164, doi:10.1016/j.bbabio.2014.03.007 (2014).

63 Dietrich, W. & Klimmek, O. The function of methyl-menaquinone-6 and polysulfide reductase membrane anchor (PsrC) in polysulfide respiration of *Wolinella succinogene*s. *Eur J Biochem* **269**, 1086-1095, doi:10.1046/j.0014-2956.2001.02662.x (2002).

64 Blumentals, II, Itoh, M., Olson, G. J. & Kelly, R. M. Role of Polysulfides in Reduction of Elemental Sulfur by the Hyperthermophilic Archaebacterium *Pyrococcus furiosus*. *Appl Environ Microbiol* **56**, 1255-1262, doi:10.1128/aem.56.5.1255-1262.1990 (1990).

65 Ma, K. & Adams, M. W. Sulfide dehydrogenase from the hyperthermophilic archaeon *Pyrococcus furiosus*: a new multifunctional enzyme involved in the reduction of elemental sulfur. *J Bacteriol* **176**, 6509-6517, doi:10.1128/jb.176.21.6509-6517.1994 (1994).

66 Zopfi, J., Ferdelman, T. G. & Fossing, H. in *In Sulfur Biogeochemistry – Past and Pre- sent.* (eds J. P. Amend, K. J. Edwards, & T. W. Lyons) pp. 97–116. (Geological Society of America, , 2004).

67 Bak, F. & Cypionka, H. A novel type of energy metabolism involving fermentation of inorganic sulphur compounds. *Nature* **326**, 891-892, doi:10.1038/326891a0 (1987).

68 Bak, F. & Pfennig, N. Chemolithotrophic growth of *Desulfovibrio sulfodismutans* sp. nov. by disproportionation of inorganic sulfur compounds. *Archives of Microbiology* **147**, 184-189, doi:10.1007/BF00415282 (1987).

69 Finster, K., Liesack, W. & Thamdrup, B. Elemental sulfur and thiosulfate disproportionation by *Desulfocapsa sulfoexigens* sp. nov., a new anaerobic bacterium isolated from marine surface sediment. *Appl Environ Microbiol* **64**, 119-125, doi:10.1128/aem.64.1.119-125.1998 (1998).

70 Jackson, B. E. & McInerney, M. J. Thiosulfate disproportionation by *Desulfotomaculum thermobenzoicum*. *Appl Environ Microbiol* **66**, 3650-3653, doi:10.1128/aem.66.8.3650-3653.2000 (2000).

71 Janssen, P. H., Schuhmann, A., Bak, F. & Liesack, W. Disproportionation of inorganic sulfur compounds by the sulfate-reducing bacterium *Desulfocapsa thiozymogenes* gen. nov., sp. nov. *Archives of Microbiology* **166**, 184-192, doi:10.1007/s002030050374 (1996).

72 Jørgensen, B. B. A thiosulfate shunt in the sulfur cycle of marine sediments. *Science* **249**, 152-154, doi:10.1126/science.249.4965.152 (1990).

73 Jørgensen, B. B. & Bak, F. Pathways and microbiology of thiosulfate transformations and sulfate reduction in a marine sediment (kattegat, denmark). *Appl Environ Microbiol* **57**, 847-856, doi:10.1128/aem.57.3.847-856.1991 (1991).

74 Krämer, M. & Cypionka, H. Sulfate formation via ATP sulfurylase in thiosulfate- and sulfite-disproportionating bacteria. *Archives of Microbiology* **151**, 232-237, doi:10.1007/BF00413135 (1989).

75 Heinzinger, N. K., Fujimoto, S. Y., Clark, M. A., Moreno, M. S. & Barrett, E. L. Sequence analysis of the phs operon in *Salmonella typhimurium* and the contribution of thiosulfate reduction to anaerobic energy metabolism. *J Bacteriol* **177**, 2813-2820, doi:10.1128/jb.177.10.2813-2820.1995 (1995).

76 Aird, B. A., Heinrikson, R. L. & Westley, J. Isolation and characterization of a prokaryotic sulfurtransferase. *J Biol Chem* **262**, 17327-17335 (1987).

77 Etchebehere, C. & Muxí, L. Thiosulfate reduction and alanine production in glucose fermentation by members of the genus *Coprothermobacter*. *Antonie Van Leeuwenhoek* **77**, 321-327, doi:10.1023/a:1002636212991 (2000).

78 Kaji, A. & Mc, E. W. Mechanism of hydrogen sulfide formation from thiosulfate. *J Bacteriol* **77**, 630-637, doi:10.1128/jb.77.5.630-637.1959 (1959).

79 Peck, H. D., Jr. & Fisher, E., Jr. The oxidation of thiosulfate and phosphorylation in extracts of *Thiobacillus thioparus*. *J Biol Chem* **237**, 190-197 (1962).

80 Ravot, G. *et al.* Thiosulfate reduction, an important physiological feature shared by members of the order Thermotogales. *Appl Environ Microbiol* **61**, 2053-2055, doi:10.1128/aem.61.5.2053-2055.1995 (1995).

81 Kawano, Y. *et al.* Improved fermentative L-cysteine overproduction by enhancing a newly identified thiosulfate assimilation pathway in *Escherichia coli*. *Appl Microbiol Biotechnol* **101**, 6879-6889, doi:10.1007/s00253-017-8420-4 (2017).

82 Finster, K. Microbiological disproportionation of inorganic sulfur compounds. *Journal of Sulfur Chemistry* **29**, 281-292, doi:10.1080/17415990802105770 (2008).

83 Frederiksen, T. M. & Finster, K. Sulfite-oxido-reductase is involved in the oxidation of sulfite in *Desulfocapsa sulfoexigens* during disproportionation of thiosulfate and elemental sulfur. *Biodegradation* **14**, 189-198, doi:10.1023/a:1024255830925 (2003).

84 Mowat, C. G. *et al.* Octaheme tetrathionate reductase is a respiratory enzyme with novel heme ligation. *Nat Struct Mol Biol* **11**, 1023-1024, doi:10.1038/nsmb827 (2004).

85 Hinojosa-Leon, M., Dubourdieu, M., Sanchez-Crispin, J. A. & Chippaux, M. Tetrathionate reductase of S*almonella thyphimurium*: a molybdenum containing enzyme. *Biochem Biophys Res Commun* **136**, 577-581, doi:10.1016/0006-291x(86)90479-1 (1986).

86 Kanao, T., Kamimura, K. & Sugio, T. Identification of a gene encoding a tetrathionate hydrolase in *Acidithiobacillus ferrooxidans*. *J Biotechnol* **132**, 16-22, doi:10.1016/j.jbiotec.2007.08.030 (2007).

87 Katoh, K., Misawa, K., Kuma, K.-i. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic acids research* **30**, 3059-3066, doi:10.1093/nar/gkf436 (2002).

88 Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution* **35**, 1547-1549,

doi:10.1093/molbev/msy096 (2018).

89 Kozlov, A. M., Darriba, D., Flouri, T., Morel, B. & Stamatakis, A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**, 4453-4455, doi:10.1093/bioinformatics/btz305 (2019).

90 Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology* **34**, 525-527, doi:10.1038/nbt.3519 (2016).

91 Graham, E. D., Heidelberg, J. F. & Tully, B. J. Potential for primary productivity in a globally-distributed bacterial phototroph. *ISME J* **12**, 1861-1866, doi:10.1038/s41396-018-0091-3 (2018).

92 Rawlings, N. D. *et al.* The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic acids research* **46**, D624-D632, doi:10.1093/nar/gkx1134 (2017).

93 Huang, L. *et al.* dbCAN-seq: a database of carbohydrate-active enzyme (CAZyme) sequence and annotation. *Nucleic acids research* **46**, D516-D521, doi:10.1093/nar/gkx894 (2017).

94 Garber, A. I. *et al.* FeGenie: A Comprehensive Tool for the Identification of Iron Genes and Iron Gene Neighborhoods in Genome and Metagenome Assemblies. *Frontiers in Microbiology* **11**, doi:10.3389/fmicb.2020.00037 (2020).

Figures

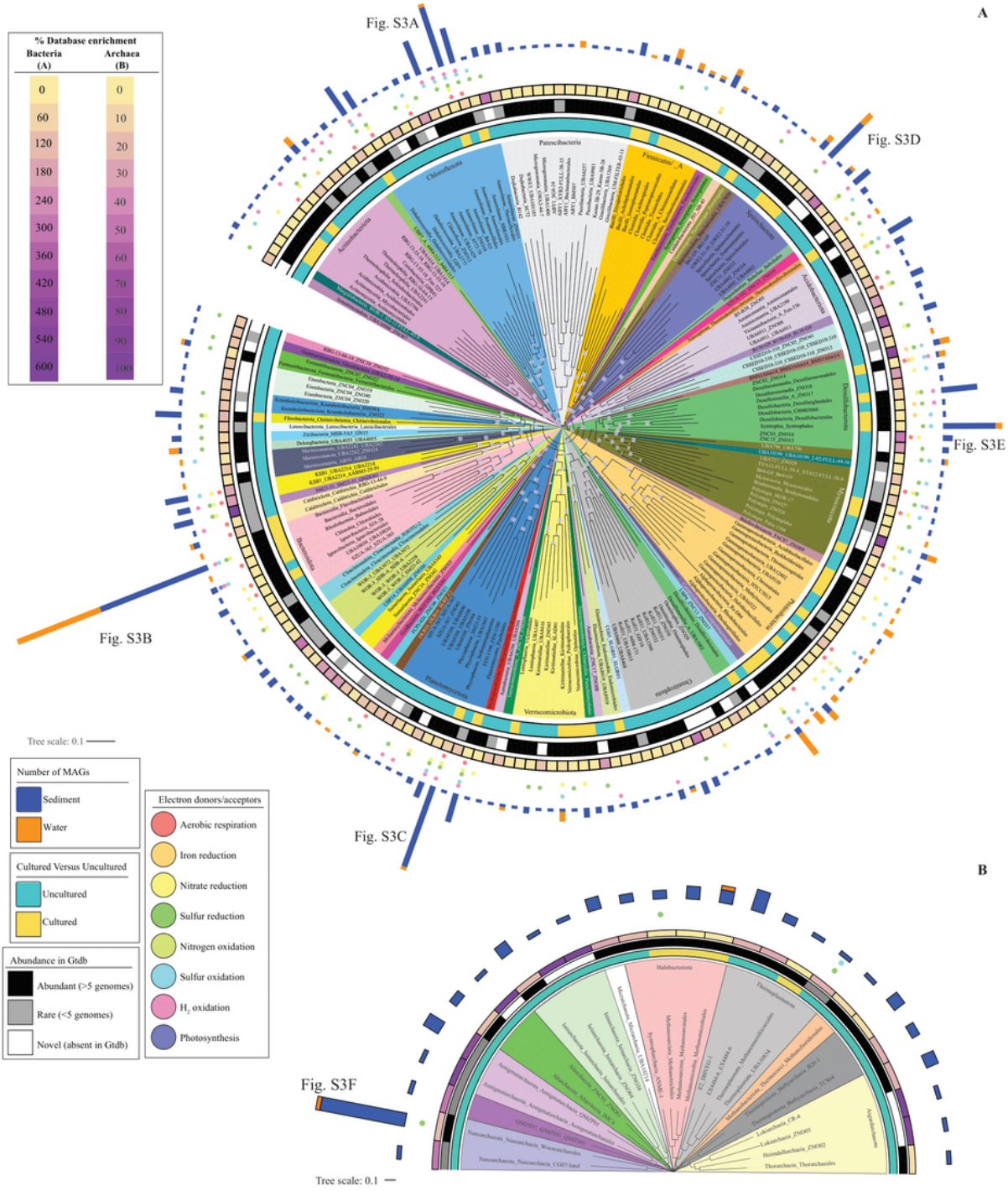

Figure 1

Phylogenomics of the 516 bacterial (A), and 114 archaeal (B) genomes analyzed in this study. The maximum likelihood trees were constructed in FastTree 31 based on the concatenated alignments of 120 (bacterial), and 122 (archaeal) housekeeping genes obtained from Gtdb-TK 32. The branches represent order-level taxonomy, are color coded by phylum, and are labeled as follows; Class_Order, phyla with > 4 orders; Phylum_Class_Order, phyla with ≤ 4 orders. Bootstrap support values are shown as bubbles for

nodes with >70% support. Tracks around the tree represent (from innermost to outermost): cultured status at the order level, abundance in Gtdb based on the number of available genomes (abundant, > 5 genomes; rare, ≤ 5 genomes; novel, no genomes in Gtdb), percentage database enrichment (number of genomes belonging to a certain order binned in the current study / number of genomes belonging to the same order in Gtdb), energy conservation capabilities depicted by colored circles as shown in the figure legend, and the number of MAGs belonging to each order binned from the sediment (blue bars) and the water (orange bars). For orders with ≥20 genomes, the family-level delineation is shown in Figure S3. ZNC = novel class; ZNO = novel order.

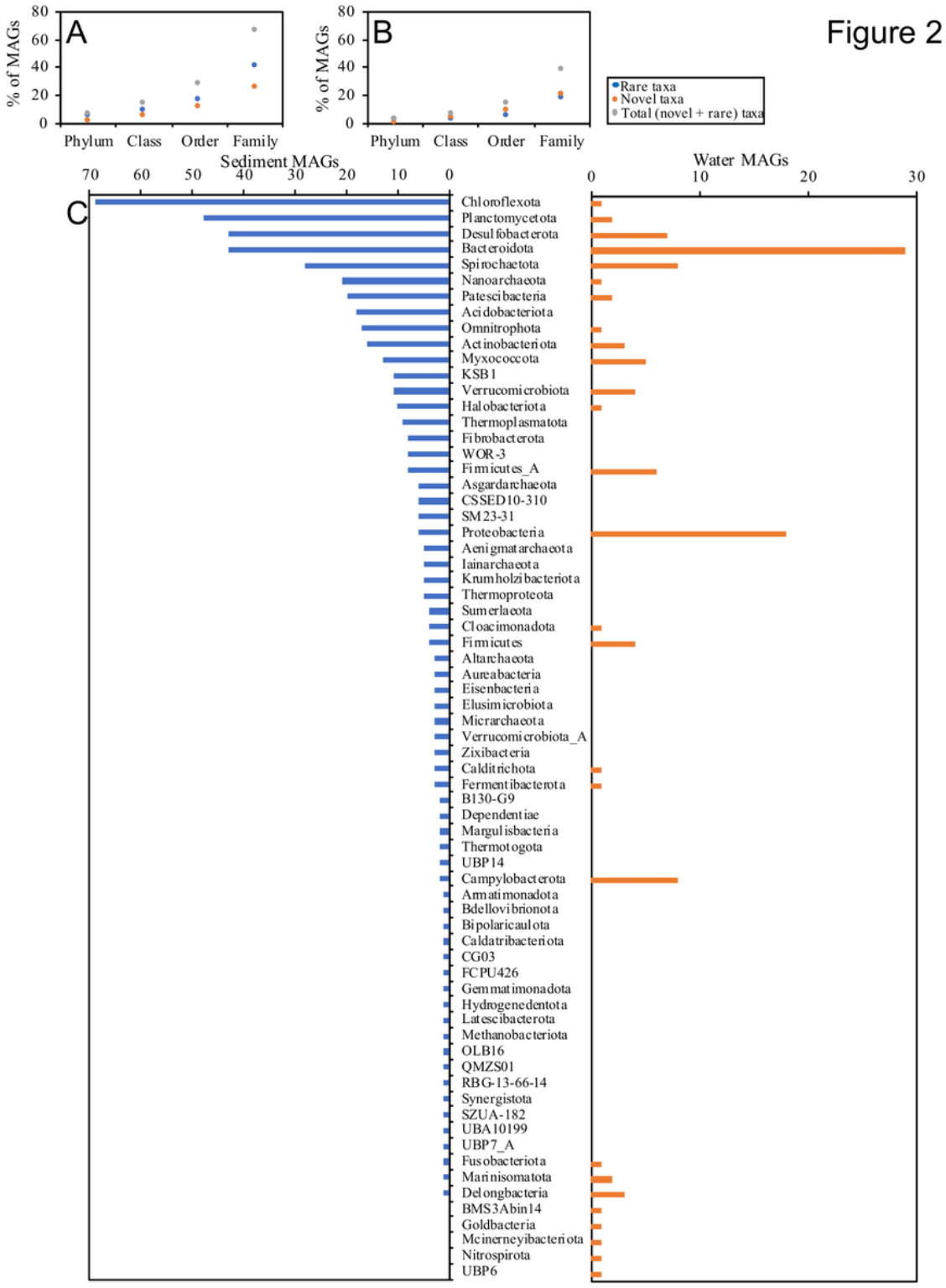

Figure 2

Novelty, rarity, and phylum-level makeup in Zodletone sediment and water communities. Genomes belonging to novel (orange), and LRD (blue) lineages are shown as a percentage of total binned genomes in the sediment (A) and the water (B) communities. The sum of novel and LRD genomes is shown in grey. (C) Phylum-level affiliation for sediment versus water genomes. Number of genomes belonging to each phylum is shown for the sediment (blue bars on the left) and the water (orange bars on the right).

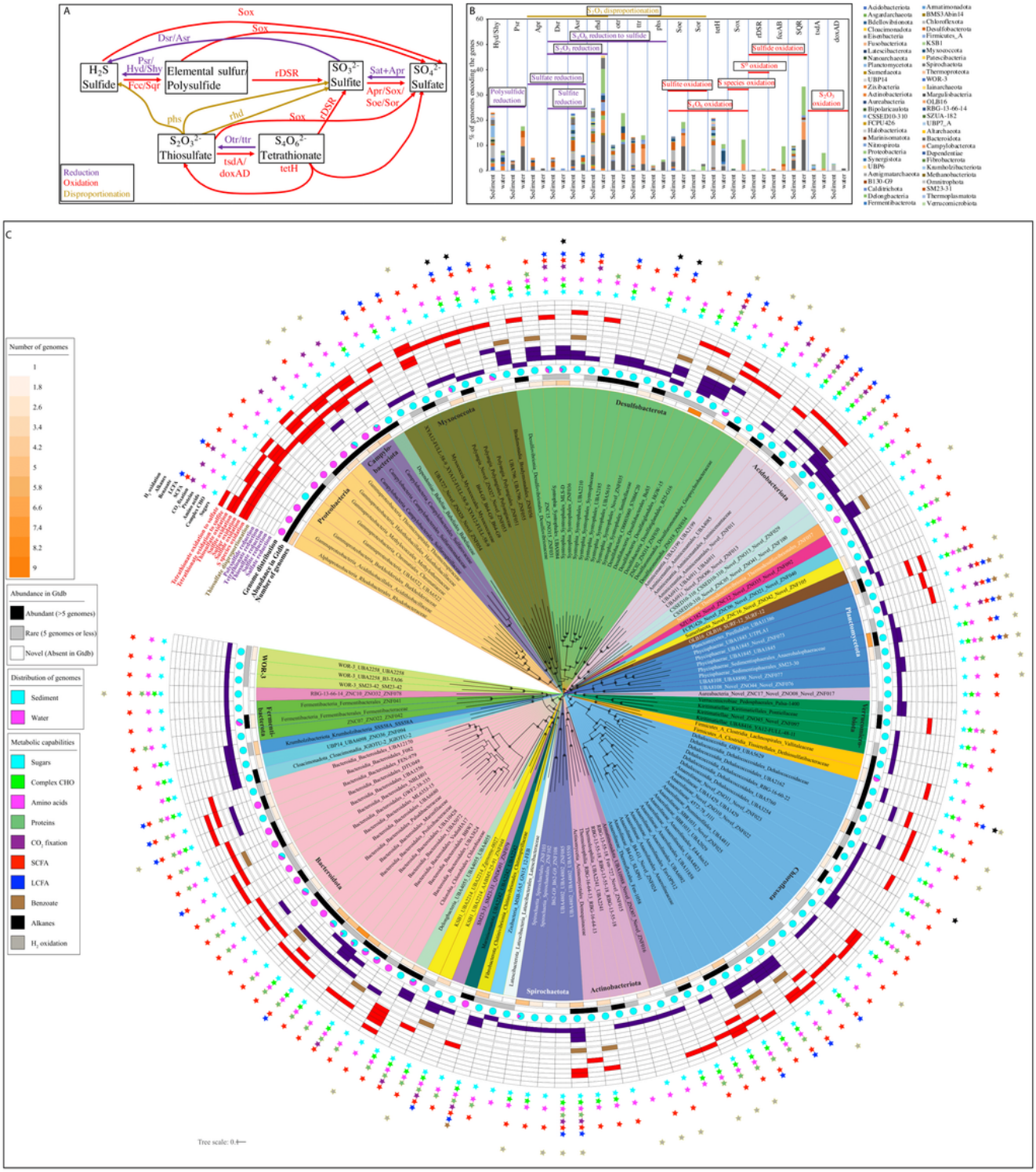

Figure 3

Sulfur cycle in Zodletone spring. (A) Diagram of sulfur transformations in the spring. Sulfur species are shown in black boxes. Purple arrows, Reduction reactions; red arrows, oxidation reactions; golden brown arrows, disproportionation reactions. Gene(s) names are shown on the arrows. (B) Phylum-level distribution of the S-cycling genes in sediment and water genomes. Horizontal bars depict processes involving more than one reaction; purple, reduction; red, oxidation; golden brown, disproportionation. (C)

Family-level taxonomy of the genomes involved in S cycling. Branches are color coded by phylum, represent family-level taxonomy, and are labeled as follows; Class_Order_Family, phyla where ≥ 3 families are involved in S cycling; Phylum_Class_Order_Family, phyla with ≥ 2 families involved in S cycling. Bootstrap support values are shown as bubbles for nodes with >70% support. Tracks around the tree represent: heatmap for the number of genomes in each family, abundance in Gtdb, distribution in the sediment (cyan) versus water (magenta), sulfur reduction pathways (5 tracks in purple), thiosulfate disproportionation pathways (1 track in golden brown), sulfur oxidation pathways (7 tracks in red), and substrates predicted to support growth depicted by colored stars (as shown in the figure legend). ZNC, novel class; ZNO, novel order; and ZNF, novel family.

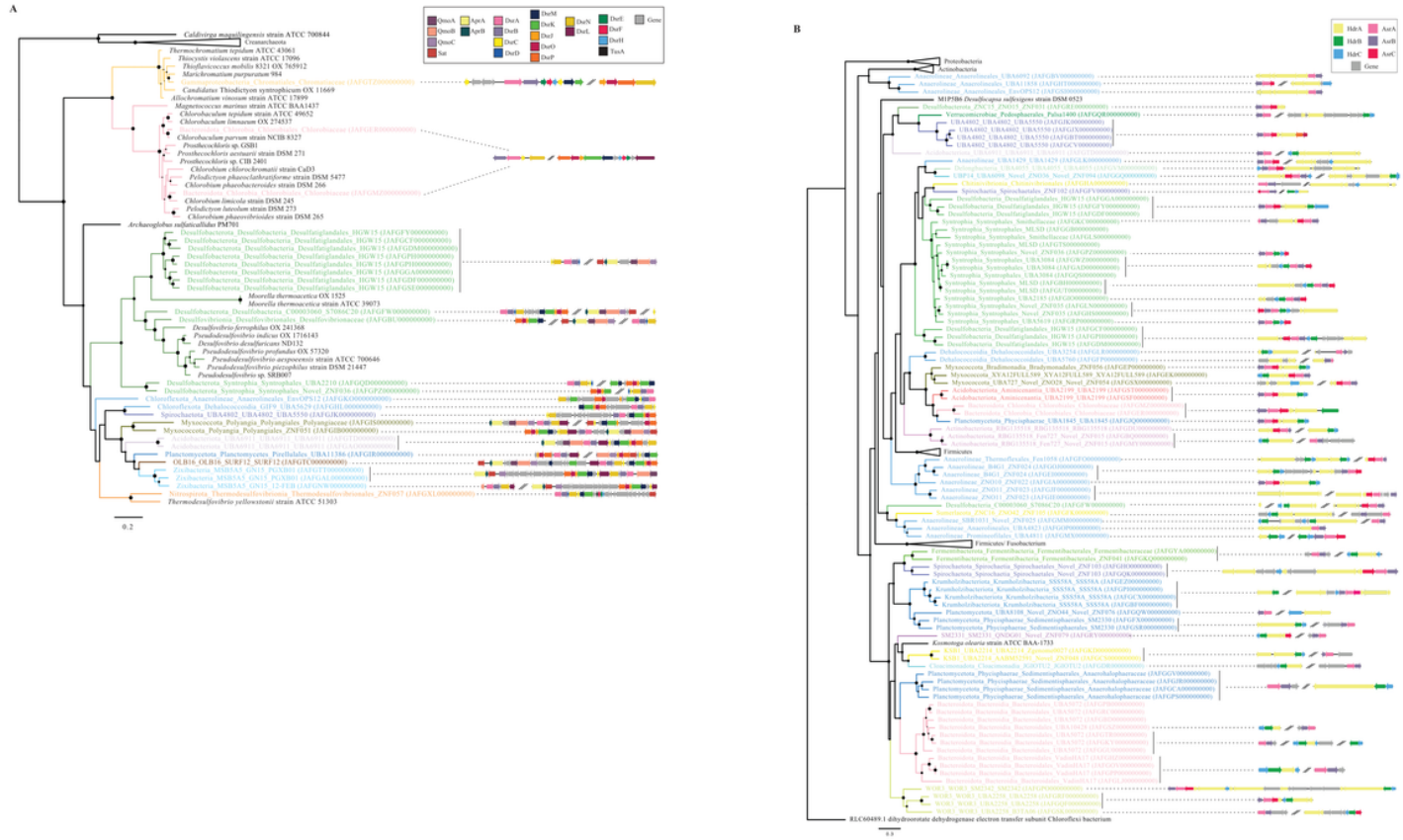

Figure 4

Phylogenetic affiliation and contig organization of selected sulfur reduction proteins. (A) Phylogeny of the dissimilatory sulfite reductase DsrAB [EC:1.8.99.5] concatenated proteins (A), anaerobic sulfite reductase subunit B AsrB (B). Alignments were created in Mafft 33 and maximum likelihood trees were constructed in RAxML 34. Bootstrap support values are shown as bubbles for nodes with >50% support. Branches and branch labels are color coded by phylum for Zodletone sequences. Branch labels depict classification to family level followed by the NCBI genome accession number. Reference sequences are shown in black with the Uniprot accession numbers. Contig organizations of the DSR and ASR loci in selected Zodletone genomes are shown to the right of the trees. Genes are color coded as shown in the top right corner. Unrelated genes are shown by grey arrows. Gene maps were created in R using the package genoplotR 35.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarytablescopy.xlsx
- SupptextNatue.docx