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

C. Ryan Hahn[1], Ibrahim F. Farag[1$], Chelsea L. Murphy[1], Mircea Podar[2,3], Mostafa S. Elshahed[1], and Noha H. Youssef[1]*

[1] Department of Microbiology and Molecular Genetics, Oklahoma State University, Stillwater, OK, USA,
[2] Department of Microbiology, University of Tennessee Knoxville, Knoxville, TN, USA, and [3] Oak Ridge National Laboratory, Oak Ridge, TN, USA

[$] Present address: University of Delaware, Newark DE.

Classification: Biological Sciences-Microbiology

***Correspondence:** Noha H. Youssef: noha@okstate.edu

Abstract

Life emerged and diversified in the absence of molecular oxygen. 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. 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. We document a remarkably diverse microbial community in the anoxic spring sediments, with 340/516 (65.89%) of genomes recovered in a metagenomic survey belonging to 200 bacterial and archaeal families that were either previously undescribed or that exhibit an extremely rare distribution on the current earth. Such diversity is underpinned by the widespread occurrence of sulfite-, thiosulfate, tetrathionate-, and sulfur-reduction, and paucity of sulfate-reduction machineries in these taxa; hence greatly expanding lineages mediating reductive sulfur cycling processes in the tree of life. Analysis of the overlaying water community demonstrated that oxygen intrusion lead to the development of a significantly less diverse community dominated by well-characterized lineages and a 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 in an anoxic earth, but have failed to adapt to earth's progressive oxygenation.

Introduction

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. Reductive processes employ sulfur oxyanions or elemental sulfur as terminal electron acceptors in anaerobic respiratory schemes linked to heterotrophic or autotrophic growth. Oxidative processes, on the other hand, employ sulfides or elemental sulfur as electron donors, powering chemolithotrophic and photosynthetic growth.

Thermodynamic considerations limit reductive sulfur processes to habitats where oxygen is limited. This is reflected in the global distribution of sulfate (SO_4^{2-}), sulfite (SO_3^{2-}), thiosulfate ($S_2O_3^{2-}$), tetrathionate ($S_4O_6^{2-}$), and elemental sulfur (S^0) reducing microorganisms (henceforth collectively referred to as SRM) in permanently and seasonally anoxic and hypoxic habitats in marine [1, 2, 3], freshwater [4], terrestrial [5], and subsurface [6] ecosystems. Sulfate is highly abundant on the current earth, and hence sulfate-reduction dominates reductive processes in the global sulfur cycle; although reduction and disproportionation of the intermediate sulfur species, e.g. sulfur [7, 8], sulfite [8], thiosulfate and tetrathionate [9, 10] could be significant in localized settings.

The history of earth's sulfur cycle is a prime example of a geological-biological feedback loop, where the evolution of biological processes is driven by, and dramatically impacts, the earth's biogeochemistry. The earth's surface was completely anoxic during the first two billion years of its history, and the availability and speciation of various sulfur species greatly differed from its current values. Sulfate levels were significantly lower when compared to current values in oceanic water (28 mM), with estimates of <200 μM-1mM from the Archean up to the Paleoproterozoic (2.3 Gya) [11, 12, 13, 14]. On the other hand, intermediate sulfur species appear to have played an important role in shaping the ancient sulfur cycle [15]. Modeling suggests that mM

levels of SO_3^{2-} were attained in the Archaean 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 [15, 16].

The evolution of life (3.8 to 4.0 Gy) in the early Archean era and the subsequent evolution of major bacterial and archaeal clades in the Archean and early Proterozoic eons [17] occurred within this background of anoxia and characteristic sulfur-chemistry. As such, it has been speculated that organisms using intermediate forms of sulfur were likely more common than sulfate-reducing organisms [15]. However, while isotopic fractionation, modeling, and microscopic studies could provide clues on prevailing sulfur speciation patterns and prevalent biological processes, the identity of microorganisms mediating such processes is unknown. This is mostly due to constrains on preservation of nucleic acids and other biological macromolecules, with the oldest successful DNA sequenced sample being only 1.2 M years old [18].

Investigation of the microbial community in modern ecosystems with conditions resembling those prevailing in the ancient earth could provide important clues to the nature, identity, and evolutionary trajectory of microorganisms that thrived under conditions prevailing prior to earth's oxygenation and the associated changes in the sulfur cycle. In Zodletone spring, a surficial anoxic spring in southwestern Oklahoma, the prevailing conditions are analogous to those predominant on the earth's surface in the late Archean/early Proterozoic eons 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 occurs 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. Further, the sediments at the source of the spring are overlaid by an air-exposed water column, where oxygen intrusion leads to a vertical oxygen gradient (from oxic in the top 1 µm, to hypoxic in the middle, to anoxic in deeper layers overlaying the sediments) that simulates oxygen intrusion into such habitats during the great oxygenation event (Figure S1). As such the spring represents a readily accessible habitat where spatial gradients remarkably correspond to temporal geological transitions. Here, we combined metagenomic, metatranscriptomic, and amplicon-based approaches to characterize the microbial community in Zodletone spring. Our results provide a glimpse of the community mediating the ancient sulfur cycle, significantly expand the overall microbial diversity by the description of a wide range of novel lineages, and greatly increase the number of lineages documented to mediate reductive sulfur processes in the microbial tree of life.

Results

Novel phylogenetic diversity in Zodletone sediments. Metagenomic sequencing of the spring sediments yielded 281 Gbp, 79.54% of which assembled into 12 Gbp contigs, with 6.8 Gbp contigs longer than 1Kbp. 1,848 genomes were binned, 683 of which passing quality control criteria, and 516 remaining after dereplication (Table S1). These MAGs represented 64 phyla or candidate phyla (53 bacterial and 11 archaeal), 127 classes, 198 orders, and 300 families (Figure 1a-b). Diversity assessment utilizing small subunit ribosomal protein S3 from assembled contigs (n=2079), as well as a complementary 16S rRNA illumina sequencing effort (n=309,074 amplicons), identified a higher number of taxa (82 phyla and 1679 species in the ribosomal protein S3 dataset, and 69 phyla and 1050 species in 16S rRNA dataset) (Figure S2). Nevertheless, the overall community composition profiles generated from all three approaches were broadly similar (Figure S2), suggesting that the MAG list broadly reflects the sediment microbial community.

Assessment of novelty and degree of uniqueness of sediment MAGs identified a remarkably high number of previously undescribed lineages (1 phylum, 14 classes, 43 orders, and 97 families), as well as Lineages exhibiting Rare global Distribution patterns compared to other present time earth environments (defined as lineages represented by 5 genomes or less in GTDB r95, henceforth LRD) (11 phyla, 24 classes, 45 orders, and 113 families) in the spring. (Figure 1, 2a, b). At the family level, 132 (25.58%), and 208 (40.03%) genomes clustered into 97 novel and 113 LRD families respectively, bringing the proportion of genomes belonging to novel or LRD families in Zodletone sediments to 65.89%. The high level of novelty in the sediment MAGs is reflected in an average RED value of 0.76, a value that is slightly lower than the median RED value for designation of a novel family (0.77) [19].

The 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 abundant phyla in Zodletone spring sediments, albeit representing only 52.52% of the total number of recovered genomes (S. Text, Figure 1, 2C). An extreme paucity of genomes belonging to the Proteobacteria (6 genomes) and Firmicutes (12 genomes), widely distributed and abundant taxa in current biomes[20], [20], and the absence of oxygen-generating Cyanobacteria (0 genomes) were observed (Figure 1, 2C). Within the Chloroflexota, 38/69 genomes belonged to 3 previously undescribed orders, 5 previously undescribed families, and multiple LRD orders and families (Figure 3a). Within the Planctomycetota, 17/47 genomes belonged to 2 previously undescribed orders, 8 previously undescribed families, and multiple LRD orders and families (Figure 3b). Within the Bacteroidota, 27/43 genomes belonged to 1 previously undescribed family, and multiple LRD families (Figure 3c). Within the Spirochaetota, 19/28 genomes belonged to one previously undescribed class, 2 previously undescribed orders, and 9 previously undescribed families, as well as multiple LRD families (Figure 3d). Within the Desulfobacterota (Figure 3e), 35/43 genomes belonged to 3 previously undescribed classes, 10 previously undescribed orders, and 7 previously undescribed families, as well as multiple LRD families. Finally, an extremely diverse community of Patescibacteria (13 different orders, 3 of which belonging to LRD orders, and 14 different families, including 6 previously undescribed and 2 LRD families), and Nanoarchaeota (2 orders and 15 families, including 5 previously undescribed and 10 LRD families) were identified in the spring sediments. A similar pattern of high proportion of previously undescribed and LRD families was identified throughout all other lineages (S. Text, Figure 1a). Therefore, in addition to expanding the number of novel lineages (classes, orders, and families), and greatly

enriching available genomes in rare, poorly represented taxa, our results highlight the uniqueness and distinction of the microbial community thriving in Zodletone spring sediments, compared to present earth environments studied so far.

Oxygen intrusion reduces the proportion of novel and rare lineages in Zodletone spring. Metagenomic sequencing of the oxygen-exposed overlaying water column community yielded 323 Gbp, 80.07% of which assembled into 3.6 Gbp contigs with 3.1 Gbp contigs >1K. 883 genomes were binned, with only 114 remaining after dereplication. Of these, 62 belonged to shared families with the sediment community, and 52 were water specific. Genomes recovered from the water column belonged to a significantly lower number of phyla (n= 27), classes (n=37), orders (n=52), and families (n=79) when compared to the euxinic sediments (Table S1). The community exhibited a much lower level of novelty and rarity at the phylum, class, order, and family levels when compared to the sediment community respectively (Figure 2 a, b). Water-specific genomes (n=52) mostly belonged to well-characterized microbial lineages e.g. families Rhodobacteraceae, and Rhodospirillaceae in the Alphaproteobacteria, families Thiomicrospiraceae, Halothiobacillaceae, Acidithiobacillaceae, Burkholderiaceae, Chromatiaceae, Methylothermaceae in the Gammaproteobacteria, families Sulfurimonadaceae, Sulfurovaceae in phylum Camplylobacterota, and well described families in the phyla Bacteroidota, Desulfobacterota (S. Text, Figure 1, Table S1). Collectively, this demonstrates a pattern where the intrusion of oxygen is associated with a negative impact on previously undescribed and LRD lineages that are prevalent in the sediment, and triggered the propagation of proliferation of cosmopolitan communities within the bacterial tree of life.

Reductive sulfur processes dominate Zodletone spring sediment communities. A total of 149 genomes (28.9 % of all genomes), belonging to 32 phyla, 51 classes, 69 orders, and 97 families

were involved in at least one reductive sulfur processes (Figure 4, Table S2). By comparison, only 21 sediment genomes (4.06% of all genomes) encoded at least one sulfur oxidation pathway (Figure 4, Table S2). The reductive sulfur-community in the spring exhibited two unique traits: First, a majority of genomes encoding such capacities belonged to novel (47 genomes) or LRD (66 genomes) lineages (Figure 4), and second: sulfite-, polysulfide-, thiosulfate-, and tetrathionate reduction appears to be more prevalent than sulfate-reduction capacities in the sediment genomes.

Sulfate-reduction capacity was observed in only 18 sediment genomes (Figure 4, 5a), but exhibited a unique community composition, when compared to well-studied marine and terrestrial habitats [1, 2, 4, 21]. Sulfate-reduction capacities were observed in mostly previously undescribed or LRD lineages within the Zixibacteria, Acidobacteriota (members of family UBA6911, equivalent to Acidobacteria group 18), Myxococcota, Bacteroidota, Planctomycetota, candidate phylum OLB16 (1 genome), as well as rare and novel lineages within the Desulfobacterota (Figure 4, S3a). Organization of the sulfate reduction genes differed between different phyla, with Myxococcota, Zixibacteria, OLB16, and Acidobacteriota genomes encoding all genes for sulfate activation and reduction, dissimilatory sulfite reduction, as well as energy conservation on one locus, while Desulfobacterota genomes encoded genes for sulfate activation and reduction as well as energy conservation on one locus with genes for dissimilatory sulfite reduction on another, as previously observed in cultured Desulfobacterota [22, 23, 24].

Sulfite (but not sulfate) reduction via the DsrAB+DsrC+DsrKMJOP system was identified in only 8 genomes belonging to 7 families within the phyla Planctomycetes, Chloroflexota, Spirochaetota, and Desulfobacterota (Figure 4, S3b). Gene organization of the *dsr* locus in the above 9 genomes differed between different phyla, with Chloroflexota, and

Spirochaetota genomes encoding all genes for dissimilatory sulfite reduction (dsrABC plus dsrKMJOP) on one locus, while Planctomycetota and Desulfobacterota genomes showed a split *dsr* locus with dsrABC on one locus and dsrMKJOP on another.

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 28 phyla, 43 (8 novel and 9 LRD) classes, 56 (18 novel, and 12 LRD) orders, and 72 (31 novel and 25 LRD) families (Figure 4, S3b), with a gene organization of the *asr* locus adjacent to the *hdr* locus in the majority of genomes. Asr-encoding genomes in the sediments included members of predominantly previously undescribed and LRD lineages within the Chloroflexota, Desulfobacterota. Planctomycetota, and Bacteroidota. The capacity was also rampant in the yet-uncultured bacterial phyla, many of which have fairly limited distribution on the current earth, e.g. the candidate phyla CSSED10-310, FCPU426, RBG-13-66-14, SM23-31, SZUA-182, UBP14, Aureabacteria, Sumerlaeota. Interestingly, all genomes belonging to the novel phylum Krumholzibacteriota, recently described from the spring sediment [25], encoded complete anaerobic sulfite reductase systems. Zodletone dissimilatory sulfite reductase (Figure S3a) and the anaerobic sulfite reductase (Figure S3b) sequences clustered with reference sequences from the same phylum, generally showing no evidence of LGT (S. text).

Sulfur (polysulfide) reduction capacities were observed in twenty Zodletone sediment genomes that encoded *psrABC* genes (Figure 4, S3c). These genomes belonged mostly to previously undescribed and LRD families within the phyla Bacteroidota, Desulfobacterota, Myxococota, Acidobacterota, Chloroflexota, and Campylobacterota (Figure 4), In addition, representatives of the cytolpasmic sulfurhydrogenase I (HydABCD system), and/or II (ShyABCD system) were

identified in 119 Zodletone sediment genomes (Figure 4). However, as described above, direct involvement of these enzymes in an ETS-associated respiration is not yet clear.

Thiosulfate disproportionation and reduction. The quinone-dependent membrane-bound molybdopterin-containing thiosulfate reductase PhsABC was encoded in 11 genomes belonging to 6 phyla, with Bacteroidota representing the major phsABC-encoding phylum (4 genomes). Within these genomes, only two (a Chloroflexota family UBA6092 genome, and a Desulfatiglandales family HGW15 genome) also encoded a dissimilatory sulfite reductase (the Asr system) akin to the Gammaproteobacteria thiosulfate disproportionating pure culture members, where the final products of thiosulfate disproportionation are expected to be only hydrogen sulfide (Figure 4). On the other hand, 5 of the 11 phsABC-encoding Zodletone genomes also encoded the sulfite dehydrogenase SoeABC system, akin to Desulfobacterota and Firmicutes pure culture members, where the final products of thiosulfate disproportionation are expected to be both hydrogen sulfide and sulfate. (Figure 4).

In addition to the phsABC system, 14 Zodletone genomes belonging to 6 phyla (Desulfobacterota,Acidobacteriota, Chloroflexota, Bacteroidota, Spirochaetota, and Myxococcota) encoded a rhodanase-like enzyme [EC: 2.8.1.1 or EC: 2.8.1.3] for thiosulfate disproportionation, as well as enzymes for both sulfite oxidation (by means of reversal of sulfate reduction via Sat+AprAB, or the sulfite dehydrogenase SoeABC), and sulfite reduction (via the dissimilatory sulfite reductases Dsr or Asr), where the final products of thiosulfate disproportionation are expected to be both hydrogen sulfide and sulfate (Figure 4).

Tetrathionate reduction. Seventy-three Zodletone sediment genomes encoded the octaheme tetrathionate reductase (OTR) enzyme. These genomes belonged to 14 phyla with major contribution from Bacteroidota (30 genomes), Chloroflexota (10 genomes), and Desulfobacterota

(10 genomes). In addition to Otr, 68 Zodletone genomes encoded the Ttr enzyme system. These genomes belonged to 14 phyla with major contribution from Chloroflexota (22 genomes), and Desulfobacterota (20 genomes). As shown previously in *Salmonella typhimurium* [26], in presence of means for thiosulfate disproportionation/reduction and sulfite reduction, the thiosulfate produced as a result of tetrathionate reduction could be further reduced to sulfide. Out of the 105 sediment genomes encoding the Otr, and/or Ttr enzymes, only 12 genomes also encoded thiosulfate and sulfite reduction enzymes. These genomes belonged to the phyla Acidobacteriota, Chloroflexota, Desulfobacterota, Myxococcota, and Spirochaetota.

Substrates supporting sulfidogenic capacities at Zodletone spring. Within lineages mediating reductive sulfur processes in Zodletone sediments (n=98), a wide range of substrates supporting sulfidogenesis were identified (Table S3, Figure 4). These included hexoses (26-87% of sulfidogenic lineages), pentoses (30-41% of sulfidogenic lineages), amino acids and peptides (39% of lineages), short chain fatty acids, e.g. lactate, propionate, butyrate, and acetate (22-73% of lineages), long chain fatty acids (29% of lineages), aromatic hydrocarbons (3% of lineages), and short chain alkanes (6% of lineages). Autotrophic capacities with hydrogen as the electron donor were identified in 28% of sulfidogenic lineages.

Transcriptomic analysis. Transcriptional expression of genes involved in S-species reduction/disproportionation was analyzed in the spring sediments (Figure 5). All S-species reduction/ disproportionation genes discussed above were identified and mapped to 51 distinct phyla. Total transcription levels of the Asr system were 4-times higher than the Dsr system, consistent with the higher number of Zodletone sediment genomes encoding the Asr system compared to the Dsr system. Transcription of Asr system genes were mapped to the phyla Chloroflexota, Planctomycetota, Desulfobacterota, Bacteroidota, Fermentibacterota,

Acidobacteriota, CSSED10-310, Actinobacteriota, Spirochaetota, WOR-3, and Fibrobacterota; while DSR genes to Desulfobacterota, Acidobacteriota, Zixibacteria, and Myxococcota (Dsr) (Figure 5). Sulfate reduction genes (Sat, AprAB, and QmoABC) were also transcribed with major contribution from Desulfobacterota, Myxococcota, Zixibacteria, and Acidobacteriota. Transcription of the thiosulfate disproportionating rhodanese-like enzyme [EC: 2.8.1.1 or EC: 2.8.1.3], thiosulfate reductase *phsABC*, tetrathionate reduction genes *ttrABC*, octaheme tetrathionate reductase *otr* , and *psrABC* for polysulfide reduction, and cytoplasmic sulfurhydrogenases I and II (*hyd/shy* systems) was also identified (Figure 5, S. text for detailed contributions of taxa).

Oxidative sulfur processes dominate Zodletone water community.

Reductive sulfur-processes were extremely sparse in the water community (S. text). In contrast oxidative sulfur processes dominated the water community, with pathways encoding sulfide, sulfur, thiosulfate, tetrathionate, and/or sulfite oxidation to sulfate present in 59/114 (51.8%) of water genomes, belonging to 13 phyla, 16 classes, 25 orders, and 43 families, respectively. The oxidative sulfur community in the water belonged to mostly well-characterized lineages (Table S3, Figure 4), with only 8 and 10 genomes involved in oxidative sulfur processes belonged to previously undescribed, and LDR families, respectively. A complete SOX system, putatively mediating oxidation of a wide range of reduced sulfur-species to sulfate was encoded in genomes belonging to well-characterized families within the Proteobacteria (11 genomes in Acidithiobacillaceae, Burkholderiaceae, Halothiobacillaceae, Rhodobacteraceae, and Thiomicrospiraceae) and Campylobacterota (3 genomes in the family Sulfurimonadaceae) (Figure 4). The capacity for sulfide oxidation to sulfur (sulfide dehydrogenase and/or the sulfide:quinone oxidoreductase Sqr) was encoded in thirty nine water genomes within the phyla

Bacteroidota, Proteobacteria, Campylobacterota, Chloroflexi, Desulfobacterota, Marinisomatota, and Nitrospirota (S. Text, Figure 4). Only two of the above thirty-nine genomes (a Proteobacteria genome and a Nitrospirota genome) encoded the capacity to further oxidize the sulfur/polysulfide to sulfite via the reversal of the Dsr system, the sulfite dehydrogenase (quinone) SoeABC, or the sulfite dehydrogenase (cytochrome) SorAB system was encountered in 1, 22, and 3 genomes respectively, belonging to well characterized families within Bacteroidetes, Proeobacteria, and Campilobacterota (S. text, Figure S4). Finally, for thiosulfate oxidation, eight water genomes (Proteobacteria and Flavobactereaceae) encoded thiosulfate to tetrathionate oxidation capacities via either the thiosulfate dehydrogenase *tsdA* [EC: 1.8.2.2] or *doxAD* [EC: 1.8.5.2]. Two of these 8 genomes (1 Rhodobacteraceae, and 1 Halothiobacillaceae genomes) also encoded tetrathionate hydrolase (*tetH*) [27] known to cleave tetrathionate to thiosulfate, sulfur, and sulfate. Simultaneous identification of the SOX system and both forms of sulfide dehydrogenase (fccAB and Sqr) imply that these two genomes encode the capacity for complete thiosulfate oxidation to sulfate.

Discussion

The microbial community in Zodletone spring sediments exhibited a high level of phylogenetic diversity, novelty, and rarity (Figure 1-3). Conversely, representatives of lineages that predominate in most present earth environments, e.g. Proteobacteria, Firmicutes, and Cyanobacteria were absent or extremely sparse within the sediments. The community in the spring sediments was also characterized by a high proportion of SRM, and the prevalence of lineages mediating sulfur cycle intermediates (sulfite, thiosulfate, tetrathionate, and elemental sulfur). Many of the organisms mediating reductive sulfur-cycling processes belonged to novel and LRD lineages, hence greatly expanding the range of SRM within the tree of life (Figure 4).

What drives the assembly, propagation, and maintenance of such a diverse, novel, and distinct community in the spring sediments? The high level of diversity, novelty, and rarity within Zodletone spring sediment SRM community could be attributed to two main factors. First, the availability of a wide range of sulfur cycle intermediates in concentrations much higher than sulfate, in contrast to sulfate-predominance in current ecosystems [2]. Such pattern selects for a more diverse community of SRM in the spring, when compared to predominantly sulfate-driven marine and freshwater ecosystems (Figure 4). Second, additional factors usually constraining SRM growth in several habitats, e.g. diel or seasonal intrusion of oxygen, Fe and NO_3 [1, 28, 29, 30], recalcitrance of available substrates [6, 31, 32], temperature [33, 34], pH [21, 35, 36], salinity [37], and pressure extremes [31, 38], or combinations thereof, are absent in the spring. Therefore, while the reductive global sulfur cycle appears to be dominated by a few sulfate-reducing lineages within the Desulfobacterota, and to a lesser extent the Firmicutes, as well as Thermodesulfobacteria and *Archaeoglobus* in high temperature habitat, the SRM community in the spring is extremely diverse, encompassing a wide range of previously undescribed and LRD lineages (Figure 4, 5,

S4-S6).

Sulfate-reducing organisms are the most prevalent component of the reductive sulfur cycle in most marine and aquatic ecosystems. Aspects of the ecology [2], physiology [39], and biochemistry [40, 41, 42] of dissimilatory sulfate reduction have been extensively investigated [43]. While not the most prevalent process, the sulfate-reducing community in Zodletone spring sediment exhibited a unique composition, with members of the Zixibacteria, Acidobacteriota, Myxococcota, Bacteroidota, Planctomycetota, and candidate phylum OLB16 constituting the major players, as well as rare and novel lineages within the Desulfobacterota (Desulfatiglandales, and Order C00003060), with scarce representation of canonical Desulfobacterota sulfate reducers (1 genome). While the identification of the dissimilatory sulfate reducing machinery in some of these lineages (e.g. Zixibacteria, Acidobacteriota, and Planctomycetota) has been shown before [44, 45, 46], these members rarely appear to be the dominant players in a single ecosystem.

Compared to sulfate-reduction, the ecology and diversity of microbial dissimilatory sulfite reduction has not been extensively studied. The biochemistry of the process has been examined in sulfate-reducers, when grown on sulfite [43], as well as in a few other dedicated sulfite reducers, e.g. *Desulfitobacterium* [47], *Salmonella* [48], *Shewanella* [49], and *Wolinella* [50]. A recent study suggested the importance and ancient nature of sulfite reduction in an extreme thermophilic environment in a limited diversity biofilm [51]. We document a plethora of microorganisms within the phyla Planctomycetes, Chloroflexota, Spirochaetota, and Desulfobacterota encoding the dissimilatory sulfite reductase DSR, as well as 72 additional families (31 novel and 25 LRD) encoding the anaerobic sulfite reductase. These organisms greatly expand the known sulfite reduction capacity within the domain Bacteria. Further, the

novelty or rarity of some of these families is a reflection of the dearth of present-time habitats that could support this mode of metabolism, once predominant on ancient earth.

The bulk of knowledge on thiosulfate reduction and or disproportionation comes from studies in pure cultures, e.g. members of the Desulfobulbaceae (e.g. *Desulfocapsa*) [52, 53] and the genera *Desulfovibrio* and *Desulfomonile* [54, 55, 56] within Desulfobacterota, the gammaproteobacterium *Pantoea agglomerans* [57], members of Thermodesulfobacteria [58] and Firmicutes [59] . Radioisotope tracing of different sulfur atoms showed a significant contribution of thiosulfate disproportionation to the sulfur cycle in marine [60], as well as freshwater sediments [61]. However, the lack of a marker gene for the process hinders ecological culture-independent studies. Similar to sulfite, the high levels, and constant generation of thiosulfate in Zodletone sediments sustains a highly diverse thiosulfate reducing (thiosulfate reductase plus a sulfite reduction complex), or disproportionating (thiosulfate reductase plus both sulfite reduction and sulfite oxidation systems) community with major contribution from novel or rare families in the Acidobacteriota, Chloroflexota, Desulfobacterota, KSB1, Myxococcota, and Spirochaetota. Finally, the extremely high levels of zero valent sulfur, available as soluble polysulfide, result in enriching the community with a plethora of polysulfide-reducing organisms.

As described above, the prevailing conditions in Zodletone spring (anaerobic, surficial, light-exposed, sulfidic, with abundance of S cycle intermediates) remarkably mimic the conditions that were prevalent in the late Archean and early Proterozoic eons, when the evolution of major bacterial and archaeal clades (based on recently refined timing estimates from [62]), with the notable exception of Proteobacteria and Cyanobacteria) occurring prior to oxygen evolution. This study infers that the microbial communities presumably thriving in surficial earth were extremely diverse, with an abundance of SRM lineages. The evolution of oxygenic

photosynthesis has led to the steady and inexorable accumulation of O_2 in Earth's atmosphere (the great oxidation event, GOE), with the rise of atmospheric O_2 to 1-5% of current levels between 2.4 and 2.1 billion years (Gyr) ago, and its accumulation to values comparable to modern values 500-600 Mya [63]. 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 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. 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 to one dependent on oxidative weathering leading to the release of huge amount of sulfate derived from the oxidation of pyrite and the dissolution of sulfate minerals [64], hitherto a minor byproduct of Archean abiotic and biotic reactions [15, 65]. Therefore, it appears that loss of niches associated with geological transformations could be one of the possible explanations for high extinction rates for microorganisms on earth, as well as the constant identification of rare, novel taxa within anaerobic settings. It is notable that phylogenetically novel branches with extremely rare distribution in earth (defined as phyla with 5 genomes or less in GTDB) have been consistently identified in anaerobic habitats.

The hypothesis that the GOE played an important role in the extinction of many novel anaerobic lineages, specifically those mediating SCI reductive processes is bolstered by our analysis of the overlaying microbial community in the spring. Here, spatial oxygen intrusion provides a mimicking effect to the relentless temporal progression of oxygenation through the Proterozoic (2.4-2.1 Gya) and Paleo-Phanerozoic (500-600 Mya). The analysis demonstrates a community transitioning from ancient novelty to modern commonality, with the proportion of

novel and rare taxa as well as lineages associated with reductive sulfur processes decreasing at the expense of propagation of organisms mediating oxidative sulfur processes.

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 wide range of novel lineages, and significantly enrich representation of a wide range 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 remarkable similarity to conditions prevailing prior to the GOE, we consider the spring an invaluable portal to investigate the community thriving on the earth's surface during these eons, and posit that GOE-precipitated the near extinction of a wide range of phylogenetically distinct oxygen-sensitive lineages and drastically altered the reductive sulfur-cycling community from sulfite, sulfur, and thiosulfate reducers to predominantly sulfate reduction in the current earth.

Materials and 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 [66]. The spring geochemistry has regularly been monitored during the last two decades [66, 67, 68] 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) [68, 69].

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 [70], tailored to provide better coverage for several under-represented microbial lineages. They included 515FY (5'GTGYCAGCMGCCGCGGTAA) [71], 515F-Cren (5' GTGKCAGCMGCCGCGGTAA, for Crenarchaeota) [72], 515F-Nano (5'GTGGCAGYCGCCRCGGKAA, for Nanoarchaeota) [72], 515F-TM7 (5' GTGCCAGCMGCCGCGGTCA for TM7/Saccharibacteria) [73] as forward mix and 805RB (5' GGACTACNVGGGTWTCTAAT) [74] and 805R-Nano (5'GGAMTACHGGGGTCTCTAAT, for Nanoarchaeota) [72] 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 [75] 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).

Alpha rarefaction curves indicated saturation of observed sequence features (ASVs) at a sequencing depth of 70-80k sequences.

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 [76]. 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 [77]. Taxonomic affiliations of (rpS3) groups were identified using Diamond Blast against the GTDB r95 database [19].

Contigs from the sediment and water assemblies were binned into draft genomes using both Metabat [78] and MaxBin2 [79]. DasTool was used to select the highest quality bins from each metagenome assembly [80]. CheckM was used for estimation of genome completeness, strain heterogeneity, and contamination [81]. 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 [82]. 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 [19], and were carried out using the classify_workflow in GTDB-Tk (v1.1.0) [83]. Phylogenomic analysis utilized the concatenated alignment of a set of 120 single-copy bacterial genes, and 122 single-copy archaeal genes [19] generated by the GTDB-Tk. Maximum-likelihood phylogenomic tree was constructed in FastTree using the default parameters [84].

Annotation and metabolic analysis. Protein-coding genes were predicted using Prodigal [85]. GhostKOALA [86] 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 [87], and the alignment was used to build an HMM profile using

hmmbuild (HMMER 3.1b2) [88]. 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) [88] 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 [87], followed by maximum likelihood phylogenetic tree construction using FastTree [84]. Only candidates clustering with reference sequences were deemed true hits and were assigned to the corresponding KO. Details on the genes examined for evidence of sulfate, sulfite, polysulfide, tetrathionate, and thiosulfate reduction, thiosulfate disproportionation, and various sulfur oxidation processes are provided in the supplementary document (S. text).

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 [89]. The DsrA and DsrB alignments were concatenated in MEGA X [90]. All alignments were used to construct maximum likelihood phylogenetic trees in RAxML [91]. The R package genoPlotR [92] 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 [93]. 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 [94]. Proteases, peptidases, and protease inhibitors were identified using Blastp against the Merops database [95], 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 [96] (downloaded from the dbCAN web server in September 2020) using hmmscan. FeGenie [97] was used to predict the presence of iron reduction and iron oxidation genes in individual bins.

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

Figure legends:

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 [84] based on the concatenated alignments of 120 (bacterial), and 122 (archaeal) housekeeping genes obtained from Gtdb-TK [83]. The branches represent order-level taxonomy and are color coded by phylum. For phyla with 4 orders or less branches are labeled as Phylum_Class_Order. For phyla with more than 4 orders, the phylum is shown at the base of the colored wedge and the branches are labeled as Class_Order. Lineages staring with ZN depict novel lineages (ZNC = novel class; ZNO = novel order). 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 (cultured versus uncultured), abundance in Gtdb based on the number of available genomes (abundant with more than 5 genomes, rare with 5 genomes or less, and novel with no genomes in Gtdb), percentage database enrichment (calculated as number of genomes belonging to a certain order binned in the current study as a percentage of the number of genomes belonging to the same order in Gtdb), energy conservation capabilities depicted by colored circles (salmon, aerobic respiration; orange, Fe^{3+} respiration; yellow, nitrate/nitrite reduction; dark green, reductive sulfur processes; lime green, nitrogen oxidation; cyan, oxidative sulfur-processes; pink, respiratory hydrogen oxidation; and purple, photosynthesis), and the number of MAGs belonging to each order binned from the sediment (blue bars) and the water (orange bars). For orders with 20 or more genomes, the family-level delineation is shown in Figure 3. These orders are: Anaerolineales (Fig. 3A), Bacteroidales (Fig. 3B), Sedimentisphaerales (Fig. 3C), Spirochaetales (Fig. 3D), Syntrophales (Fig. 3E), and Woesearchaeales (Fig. 3F).

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. Family-level delineation for orders with 20 or more genomes. The maximum likelihood trees were constructed in FastTree [84] based on the concatenated alignments of 120-, and 122- single copy genes, respectively, obtained from Gtdb-TK [83]. Bootstrap support values are shown as bubbles for nodes with >70% support. Families are color-coded. To the right of the trees, tracks are shown for cultured status at the family level (cultured versus uncultured), and abundance in Gtdb based on the number of available genomes (abundant with more than 5 genomes, rare with 5 genomes or less, and novel with no genomes in Gtdb).

Figure 4. Sulfur cycle in Zodletone spring. (A) Diagram of sulfur transformations predicted to take place in the spring. Different sulfur species are shown in black boxes. Reduction reactions are depicted by purple arrows, oxidation reactions are depicted by red arrows, while disproportionation reactions are depicted by golden brown arrows. The gene(s) names are shown on the arrows. (B) Phylum-level distribution of the S-cycling genes shown at the top of the figure in sediment and water genomes. Processes involving more than one gene are highlighted by horizontal bars and are color coded by reduction (purple), oxidation (red), or disproportionation (golden brown), with the name of the process shown on top of the horizontal bar. (C) Family-

level distribution of the genomes involved in S cycling in the spring. The maximum likelihood tree was constructed in FastTree [84] based on the concatenated alignments of 120 single-copy genes obtained from Gtdb-TK [83]. The branches represent family-level taxonomy and are color coded by phylum. For phyla with 2 families or less involved in S cycling, branches are labeled as Phylum_Class_Order_Family. For phyla where 3 or more families are involved in S cycling, the phylum is shown at the base of the colored wedge and the branches are labeled as Class_Order_Family. Lineages staring with ZN depict novel lineages as follows: ZNC, novel class; ZNO, novel order; and ZNF, novel family. Bootstrap support values are shown as bubbles for nodes with >70% support. Tracks around the tree represent (from innermost to outermost): heatmap for the number of genomes in each family, abundance in Gtdb based on the number of available genomes (abundant with more than 5 genomes, rare with 5 genomes or less, and novel with no genomes in Gtdb), pie charts of the breakdown of the number of genomes 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 (cyan, sugars; lime green, complex carbohydrates; magenta, amino acids; orange, proteins; purple, CO2 fixation; red, short-chain fatty acids (SCFA); blue, beta oxidation of long-chain fatty acids; brown, anaerobic benzoate/aromatic hydrocarbon degradation; black, anaerobic alkane degradation; and grey, Hydrogen oxidation).

Figure 5. Phylum-level distribution for transcribed sulfur cycling genes in Zodletone spring sediment. RNA-seq reads were pseudo-aligned to the S-cycling genes predicted in Zodletone genomes to detect exact matches using Kallisto [93]. The transcripts per million are shown on the Y-axis for the gene/ group of genes depicted at the top of the figure.

References

1. Jørgensen BB, Findlay AJ, Pellerin A. The biogeochemical sulfur-cycle in marine sediments. *Front Microbiol* **10**, 849 (2019).

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

3. van Vliet DM, *et al*. The bacterial sulfur cycle in expanding dysoxic and euxinic marine waters. *Environ Microbiol* doi.org/10.1111/1462-2920.15265, (2020).

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

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

6. Gieg LM, Davidova IA, Duncan KE, Suflita JM. Methanogenesis, sulfate reduction and crude oil biodegradation in hot Alaskan oilfields. *Environ Microbiol* **12**, 3074-3086 (2010).

7. Steudel R. Mechanism for the formation of elemental sulfur from aqueous sulfide in chemical and microbiological desulfurization processes. *Ind Eng Chem Res* **35**, 1417–1423 (1996).

8. Jelen B, Giovannelli D, Falkowski PG, Vetriani C. Elemental sulfur reduction in the deep - sea vent thermophile, *Thermovibrio ammonificans*. *Environ Microbiol* **20**, 2301-2316 (2018).

9. Zopfi J, Ferdelman TG, Fossing H. Distribution and fate of sulfur intermediates - sulfite, tetrathionate, thiosulfate, and elemental sulfur in marine sediments. In: *In Sulfur Biogeochemistry – Past and Pre- sent.* (eds Amend JP, Edwards KJ, Lyons TW). Geological Society of America, (2004).

10. Mandal M, *et al*. Cryptic roles of tetrathionate in the sulfur cycle of marine sediments: microbial drivers and indicators. *Biogeosciences* **17**, (2020).

11. Canfield DE, Habicht KS, Thamdrup B. The Archean sulfur cycle and the early history of atmospheric oxygen. *Science* **299**, 658-661 (2000).

12. Ranjan S, Todd ZR, Sutherland JD, Sasselov DD. Sulfidic anion concentrations on early earth for surficial origins-of-Life chemistry. *Astrobiology* **18**, 1023-1041 (2018).

13. Canfield DE, Raiswell R. The evolution of the sulfur-cycle. *Am J Sci* **299**, 697-723 (1999).

14. Habicht KS, M. Gade, Thamdrup B, Berg P, Canfield DE. Calibration of sulfate levels in the Archean ocean. *Science* **298**, 2372-2374 (2002).

15. Habicht KS, Canfield DE, Rethmeier J. Sulfur isotope fractionation during bacterial reduction and disproportionation of thiosulfate and sulfite. *Geochim Cosmochim Acta* **62**, 2585-2595 (1998).

16. Philippot P, Zuilen MV, Lepot K, Thomazo C, Farquhar J, Kranendonk MJV. Early archaean microorganisms preferred elemental sulfur, not sulfate. *Science* **317**, 1534-1537 (2007).

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

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

19. Parks DH, Chuvochina M, Chaumeil P-A, Rinke C, Mussig AJ, Hugenholtz P. A complete domain-to-species taxonomy for Bacteria and Archaea. *Nat Biotechnol***38**, 1079-1086 (2020).

20. Nayfach S, *et al.* A genomic catalog of Earth's microbiomes. *Nat Biotechnol*, (2020).

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

22. Zecchin S, *et al*. Rice paddy *Nitrospirae* carry and express genes related to sulfate respiration: Proposal of the new genus *Sulfobium*". *Appl Environ Microbiol* **84**, e02224-02217 (2018).

23. Venceslau SS, Stockdreher Y, Dahl C, Pereira IAC. The "bacterial heterodisulfide" DsrC is a key protein in dissimilatory sulfur metabolism. *Biochim Biophys Acta* **1837**, 1148-1164 (2014).

24. Mussmann M, *et al*. Clustered genes related to sulfate respiration in uncultured prokaryotes support the theory of their concomitant horizontal transfer. *J Bacteriol* **187**, 7126-7137 (2005).

25. Youssef NH, *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. *Syst Appl Microbiol* **42**, 85-93 (2019).

26. Price-Carter M, Tingey J, Bobik TA, Roth JR. The alternative electron a*cceptor tetrathionate supports B12-dependent anaerobic growth of* Salmonella enterica serovar *typhimurium* on ethanolamine or 1,2-propanediol. *J Bacteriol* **183**, 2463-2475 (2001).

27. Kanao T, Kamimura K, Sugio T. Identification of a gene encoding a tetrathionate hydrolase in *Acidithiobacillus ferrooxidans*. *J Biotechnol* **132**, 16-22 (2007).

28. 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).

29. Glombitza C, Egger M, Røy H, Jørgensen BB. Controls on volatile fatty acid concentrations in marine sediments (Baltic Sea). *Geochim Cosmochim Acta*, 226-241 (2019).

30. Morrison JM, Baker KD, Zamor RM, Nikolai S, Elshhaed MS, Youssef NH. Spatiotemporal analysis of microbial community dynamics during seasonal stratification events in a freshwater lake (Grand Lake, OK, USA). *PLOS ONE* **12**, e0177488 (2017).

31. Daly RA, *et al.* Microbial metabolisms in a 2.5-km-deep ecosystem created by hydraulic fracturing in shales. *Nat Microbiol* **1**, 16146 (2016).

32. Jørgensen BB, Marshall IPG. Slow microbial life in tthe seabed. *Ann Rev Mar Sci* **8**, 311-332 (2016).

33. Jørgensen BB, Isaksen MF, Jannasch HW. Bacterial sulfate reduction above 100°C in deep-sea hydrothermal vent sediments. *Science* **1756-1757**, (1992).

34. Scholze C, Jørgensen BB, Røy H. Psychrophilic properties of sulfate‐reducing bacteria in Arctic marine sediments. *Limnol Oceanog* https://doi.org/10.1002/lno.11586, (2020).

35. Sánchez-Andrea I, Sanz JL, Bijmans MF, Stams AJ. Sulfate reduction at low pH to remediate acid mine drainage. *J Hazard Mater* **269**, 98-109 (2014).

36. Sorokin DY, Kuenen JG, Muyzer G. The microbial sulfur cycle at extremely haloalkaline conditions of soda lakes. *Front Microbiol* **2**, 44 (2011).

37. 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).

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

39. Rabus R, Venceslau SS, Wöhlbrand L, Voordouw G, Wall JD, Pereira IA. A post-genomic view of the ecophysiology, catabolism and biotechnological relevance of sulphate-reducing prokaryotes. *Adv Microb Physiol* **66**, 55-321 (2015).

40. Bertran E, *et al.* Deconstructing the dissimilatory sulfate reduction pathway: Isotope fractionation of a mutant unable of growth on sulfate. *Front Microbiol* **9**, 3110 (2018).

41. Peck HD, Jr., LeGall J. Biochemistry of dissimilatory sulphate reduction. *Philos Trans R Soc Lond B Biol Sci* **298**, 443-466 (1982).

42. Wenk CB, Wing BA, Halevy I. Electron carriers in microbial sulfate reduction inferred from experimental and environmental sulfur isotope fractionations. *ISME J* **12**, 495-507 (2018).

43. Rabus R, Hansen TA, Widdel F. Dissimilatory sulfate- and sulfur-reducing prokaryotes. In: *The Prokaryotes: Prokaryotic physiology and biochemistry* (eds Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F). Springer Berlin Heidelberg (2013).

44. Anantharaman K, *et al.* Expanded diversity of microbial groups that shape the dissimilatory sulfur cycle. *ISME J* **12**, 1715-1728 (2018).

45. Hausmann B, *et al.* Peatland Acidobacteria with a dissimilatory sulfur metabolism. *ISME J* **12**, 1729-1742 (2018).

46. Flieder M, *et al.* Novel taxa of Acidobacteriota implicated in seafloor sulfur cycling. *ISME J*, doi: 10.1038/s41396-021-00992-0 (2021).

47. Villemur R, Lanthier M, Beaudet R, Lépine F. The *Desulfitobacterium* genus. *FEMS Microbiol Rev* **30**, 706-733 (2006).

48. Huang CJ, Barrett EL. Sequence analysis and expression of the *Salmonella typhimurium asr* operon encoding production of hydrogen sulfide from sulfite. *J Bacteriol* **173**, 1544-1553 (1991).

49. Shirodkar S, Reed S, Romine M, Saffarini D. The octahaem SirA catalyses dissimilatory sulfite reduction in *Shewanella oneidensis* MR-1. *Environ Microbiol* **13**, 108-115 (2011).

50. Kern M, Klotz MG, Simon J. The Wolinella succinogenes mcc gene cluster encodes an unconventional respiratory sulphite reduction system. *Mol Microbiol* **82**, 1515-1530 (2011).

51. Colman DR, Lindsay MR, Amenabar MJ, Fernandes-Martins MC, Roden ER, Boyd ES. Phylogenomic analysis of novel Diaforarchaea is consistent with sulfite but not sulfate reduction in volcanic environments on early Earth. *ISME J* **14**, 1316-1331 (2020).

52. 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 (1998).

53. Frederiksen TM, 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 (2003).

54. Sun B, Cole JR, Tiedje JM. *Desulfomonile limimaris* sp. nov., an anaerobic dehalogenating bacterium from marine sediments. *Int J Syst Evol Microbiol* **51**, 365-371 (2001).

55. Townsend GT, Suflita JM. Influence of sulfur oxyanions on reductive dehalogenation activities in *Desulfomonile tiedjei*. *Appl Environ Microbiol* **63**, 3594-3599 (1997).

56. Haschke RH, Campbell LL. Thiosulfate reductase of *Desulfovibrio vulgaris. J Bacteriol* **106**, 603-607 (1971).

57. Obraztsova AY, Francis CA, Tebo BM. Sulfur disproportionation by the facultative anaerobe *Pantoea agglomerans* SP1 as a mechanism for chromium(VI) reduction. *Geomicrobiol J* **19**, 121-132 (2002).

58. Mardanov AV, Beletsky AV, Kadnikov VV, Slobodkin AI, Ravin NV. Genome analysis of *Thermosulfurimonas dismutans*, the first thermophilic sulfur-disproportionating bacterium of the phylum Thermodesulfobacteria. *Front Microbiol* **7**, 950 (2016).

59. Jackson BE, McInerney MJ. Thiosulfate disproportionation by *Desulfotomaculum thermobenzoicum*. *Appl Environ Microbiol* **66**, 3650-3653 (2000).

60. Jørgensen BB. A thiosulfate shunt in the sulfur cycle of marine sediments. *Science* **249**, 152-154 (1990).

61. Jørgensen BB. The sulfur cycle of freshwater sediments: Role of thiosulfate. *Limnol Oceanogr* **35**, 1329-1342 (1990).

62. Marin J, Battistuzzi FU, Brown AC, Hedges SB. The timetree of prokaryotes: New insights into their evolution and speciation. *Mol Biol Evol* **34**, 437-446 (2017).

63. Lyons TW, Reinhard CT, Planavsky NJ. The rise of oxygen in Earth's early ocean and atmosphere. *Nature* **506**, 307-315 (2014).

64. Killingsworth BA, Sansjofre P, Philippot P, Cartigny P, Thomazo C, Lalonde V. Constraining the rise of oxygen with oxygen isotopes. *Nature Comm* **10**, 4924 (2019).

65. Holland HD. 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).

66. Buhring SI, *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).

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

68. Senko JM, Campbell BS, Henricksen JR, Elshahed MS, Dewers TA, Krumholz LR. Barite deposition mediated by photootrophic sulfide-oxidizing bacteria. *Geochim Cosmochim Acta* **68**, 773-780 (2004).

69. Spain AM, Najar FZ, Krumholz LR, Elshahed MS. Metatranscriptomic analysis of a high-sulfide aquatic spring reveals insights into sulfur cycling and unexpected aerobic metabolism. *Peer J*. **3**, e1259 (2015).

70. Caporaso JG, *et al.* Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci* **108**, 4516-4522 (2011).

71. Parada AE, Needham DM, Fuhrman JA. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ Microbiol* **18**, 1403-1414 (2016).

72. Podar PT, Yang Z, BjˆrnsdÛttir SH, Podar M. Comparative analysis of microbial diversity across temperature gradients in hot springs from Yellowstone and Iceland. *Front Microbiol* **11**, (2020).

73. Cross KL, *et al.* Targeted isolation and cultivation of uncultivated bacteria by reverse genomics. *Nat Biotechnol* **37**, 1314-1321 (2019).

74. 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 (2016).

75. Bolyen E, *et al.* Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* **37**, 852-857 (2019).

76. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)* **30**, 2114-2120 (2014).

77. Diamond S, *et al.* Mediterranean grassland soil C–N compound turnover is dependent on rainfall and depth, and is mediated by genomically divergent microorganisms. *Nat Microbiol* **4**, 1356-1367 (2019).

78. Kang DD, *et al.* MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359-e7359 (2019).

79. Wu Y-W, Simmons BA, Singer SW. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* **32**, 605-607 (2015).

80. Sieber CMK, *et al.* Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat Microbiol* **3**, 836-843 (2018).

81. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* **25**, 1043-1055 (2015).

82. Parks DH, *et al*. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat Microbiol* **2**, 1533-1542 (2017).

83. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* **36**, 1925-1927 (2019).

84. Price MN, Dehal PS, Arkin AP. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLOS ONE* **5**, e9490 (2010).

85. Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119-119 (2010).

86. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J Mol Biol* **428**, 726-731 (2016).

87. Sievers F, *et al*. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539-539 (2011).

88. Mistry J, Finn RD, Eddy SR, Bateman A, Punta M. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res* **41**, e121 (2013).

89. 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 Res* **30**, 3059-3066 (2002).

90. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol* **35**, 1547-1549 (2018).

91. Kozlov AM, 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 (2019).

92. Guy L, Kultima JR, Andersson SGE. genoPlotR: comparative gene and genome visualization in R. *Bioinformatics (Oxford, England)* **26**, 2334-2335 (2010).

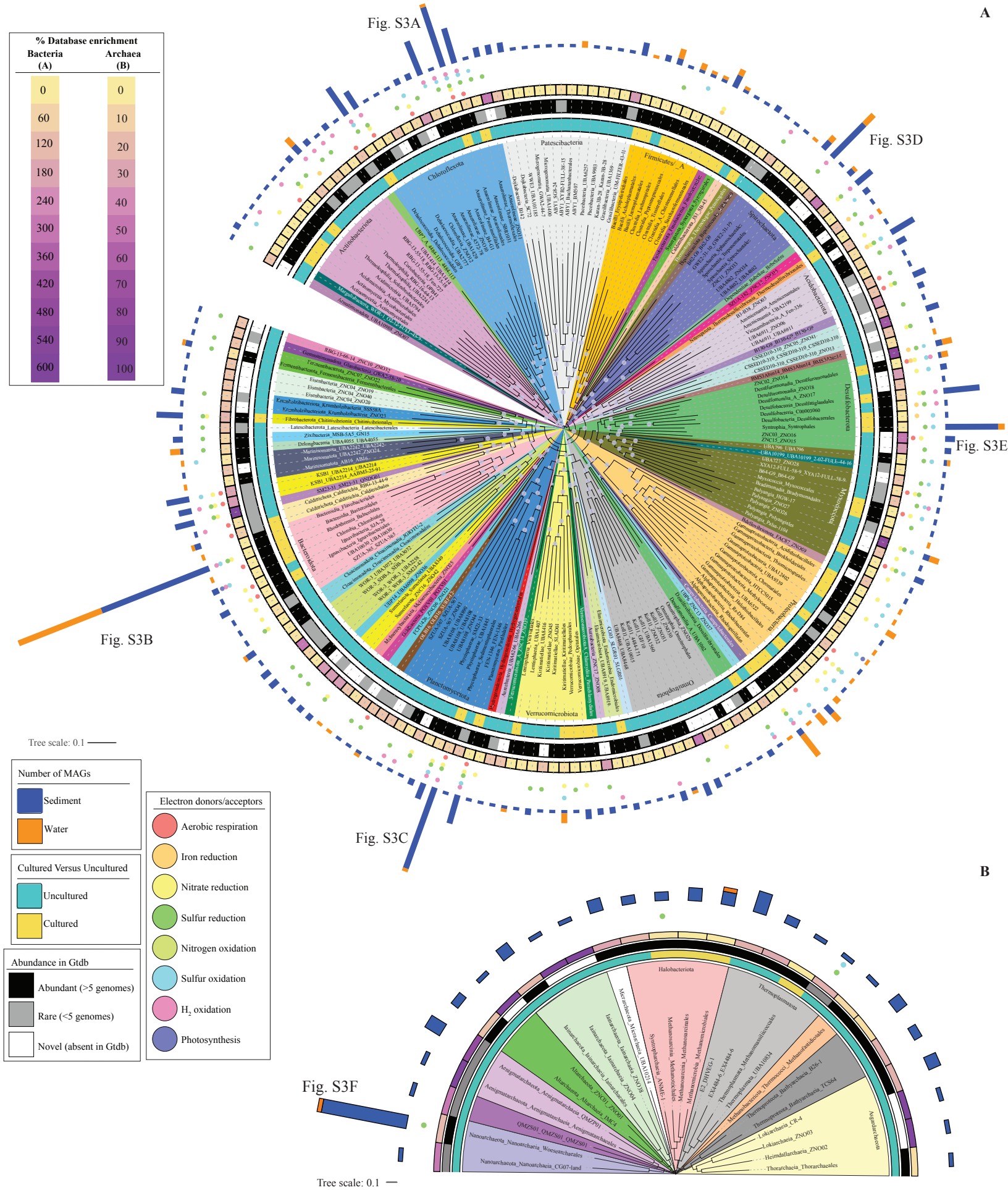
93. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* **34**, 525-527 (2016).

94. Graham ED, Heidelberg JF, Tully BJ. Potential for primary productivity in a globally-distributed bacterial phototroph. *ISME J* **12**, 1861-1866 (2018).

95. Rawlings ND, Barrett AJ, Thomas PD, Huang X, Bateman A, Finn RD. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res* **46**, D624-D632 (2017).

96. Huang L, *et al*. dbCAN-seq: a database of carbohydrate-active enzyme (CAZyme) sequence and annotation. *Nucleic Acids Res* **46**, D516-D521 (2017).

97. Garber AI, *et al*. FeGenie: A comprehensive tool for the identification of iron genes and iron gene neighborhoods in genome and metagenome assemblies. *Front Microbiol* **11**, (2020).


A
Fig. S3A
Fig. S3B
Fig. S3C
Fig. S3D
Fig. S3E
% Database enrichment
Bacteria (A)
Archaea (B)
0
60
120
180
240
300
360
420
480
540
600
0
10
20
30
40
50
60
70
80
90
100
Patescibacteria
Chloroflexota
Firmicutes_A
Spirochaetota
Acidobacteriota
Desulfobacterota
Myxococcota
Bacteroidota
Planctomycetota
Verrucomicrobiota
Omnitrophota
Tree scale: 0.1
Number of MAGs
Sediment
Water
Cultured Versus Uncultured
Uncultured
Cultured
Abundance in Gtdb
Abundant (>5 genomes)
Rare (<5 genomes)
Novel (absent in Gtdb)
Electron donors/acceptors
Aerobic respiration
Iron reduction
Nitrate reduction
Sulfur reduction
Nitrogen oxidation
Sulfur oxidation
H2 oxidation
Photosynthesis
B
Fig. S3F
Halobacteriota
Methanosarcinales_Methanosaetaceae
Methanosarcinales_Methanosarcinaceae
Methanomicrobia_Methanomicrobiales
Methanomicrobia_Methanocellales
Methanomicrobia_Methanosalinales
Methanomicrobia_Methanosphaerales
Methanomicrobia_Methanomassiliicoccales
Methanomicrobia_Methanomicrobiales
Methanomicrobia_Methanomicrobiales
Methanomicrobia_Methanomicrobi

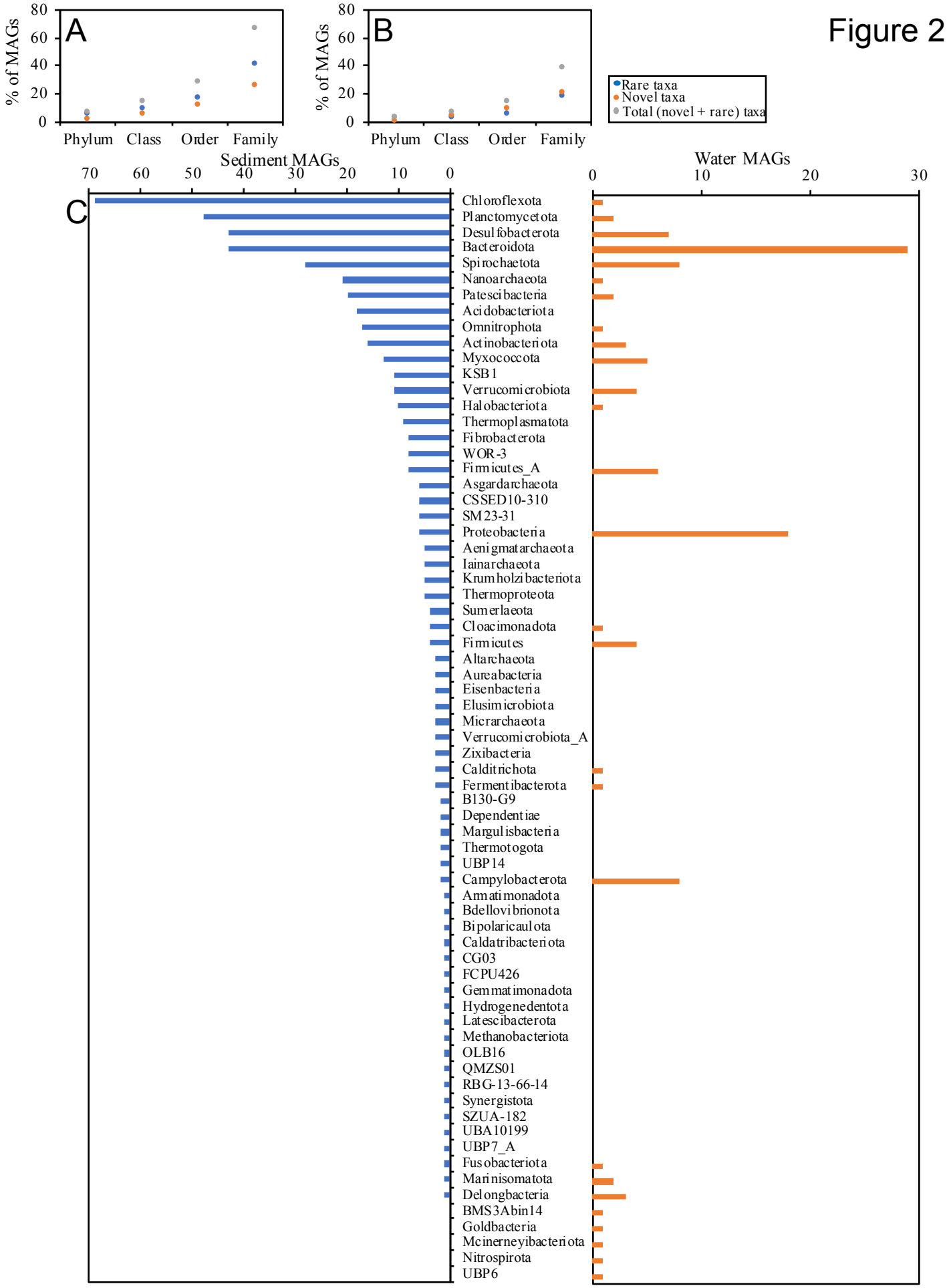
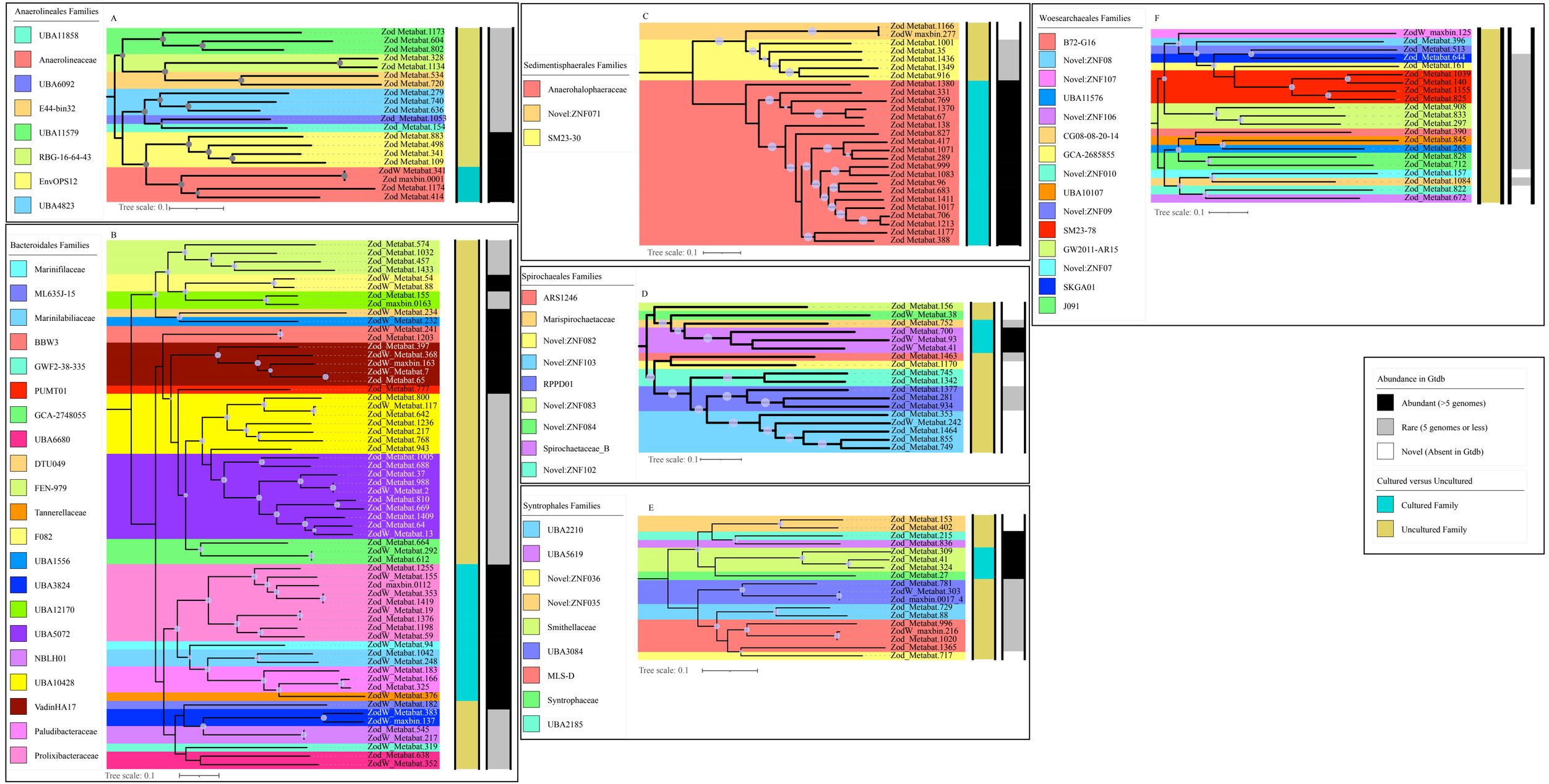
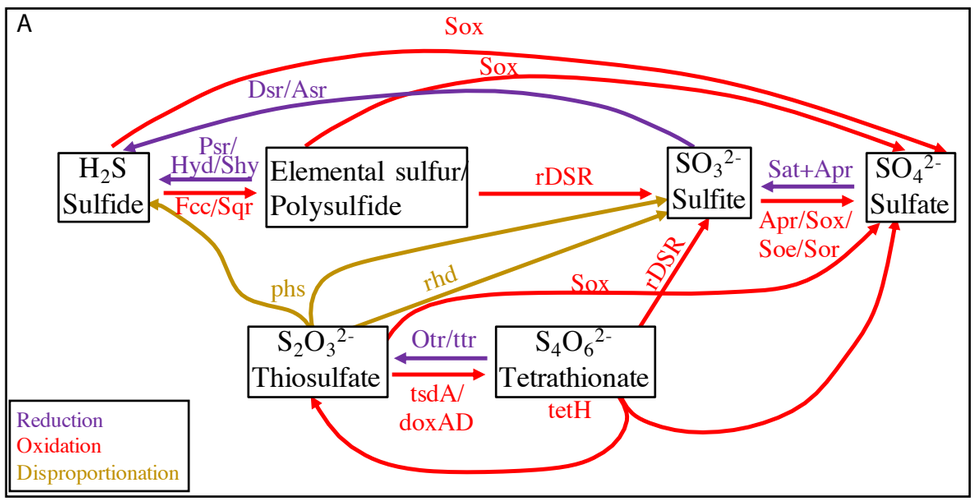
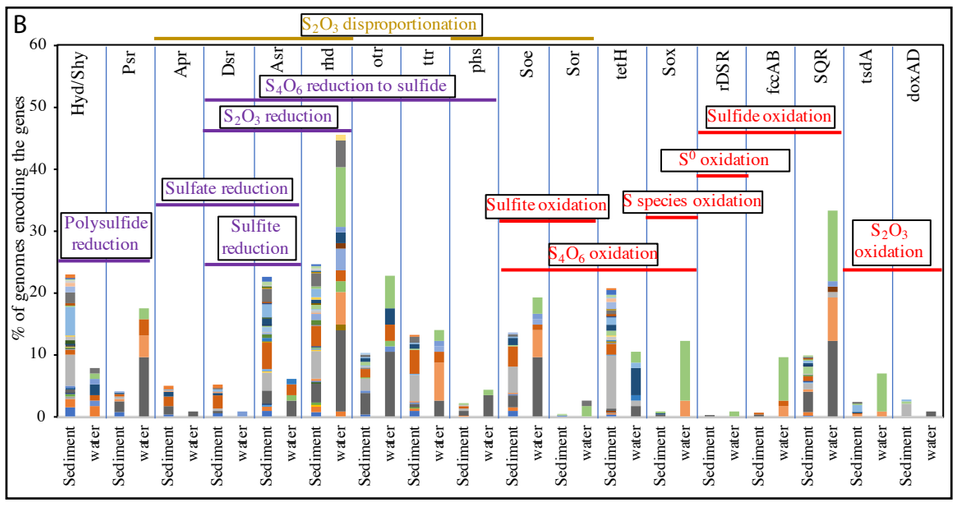


Figure 2
A
B
C
% of MAGs
80
60
40
20
0
Phylum
Class
Order
Family
Rare taxa
Novel taxa
Total (novel + rare) taxa
Sediment MAGs
70
60
50
40
30
20
10
0
Water MAGs
0
10
20
30
Chloroflexota
Planctomycetota
Desulfobacterota
Bacteroidota
Spirochaetota
Nanoarchaeota
Patescibacteria
Acidobacteriota
Omnitrophota
Actinobacteriota
Myxococcota
KSB1
Verrucomicrobiota
Halobacteriota
Thermoplasmatota
Fibrobacterota
WOR-3
Firmicutes_A
Asgardarchaeota
CSSED10-310
SM23-31
Proteobacteria
Aenigmatarchaeota
Iainarchaeota
Krumholzibacteriota
Thermoproteota
Sumerlaeota
Cloacimonadota
Firmicutes
Altarchaeota
Aureabacteria
Eisenbacteria
Elusimicrobiota
Micrarchaeota
Verrucomicrobiota_A
Zixibacteria
Calditrichota
Fermentibacterota
B130-G9
Dependentiae
Margulisbacteria
Thermotogota
UBP14
Campylobacterota
Armatimonadota
Bdellovibrionota
Bipolaricaulota
Caldatribacteriota
CG03
FCPU426
Gemmatimonadota
Hydrogenedentota
Latescibacterota
Methanobacteriota
OLB16
QMZS01
RBG-13-66-14
Synergistota
SZUA-182
UBA10199
UBP7_A
Fusobacteriota
Marinisomatota
Delongbacteria
BMS3Abin14
Goldbacteria
Mcinerneyibacteriota
Nitrospirota
UBP6


Anaerolineales Families
UBA11858
Anaerolineaceae
UBA6092
E44-bin32
UBA11579
RBG-16-64-43
EnvOPS12
UBA4823
A
Zod Metabat.1173
Zod Metabat.604
Zod Metabat.802
Zod Metabat.328
Zod Metabat.1134
Zod Metabat.534
Zod Metabat.720
Zod Metabat.279
Zod Metabat.740
Zod Metabat.636
Zod_Metabat.1053
Zod_Metabat.154
Zod Metabat.883
Zod Metabat.498
Zod Metabat.341
Zod Metabat.109
ZodW Metabat.341
Zod maxbin.0001
Zod Metabat.1174
Zod Metabat.414
Tree scale: 0.1
Bacteroidales Families
Marinifilaceae
ML635J-15
Marinilabiliaceae
BBW3
GWF2-38-335
PUMT01
GCA-2748055
UBA6680
DTU049
FEN-979
Tannerellaceae
F082
UBA1556
UBA3824
UBA12170
UBA5072
NBLH01
UBA10428
VadinHA17
Paludibacteraceae
Prolixibacteraceae
B
Zod_Metabat.574
Zod_Metabat.1032
Zod_Metabat.457
Zod_Metabat.1433
ZodW_Metabat.54
ZodW_Metabat.88
Zod_Metabat.155
Zod_maxbin.0163
ZodW_Metabat.234
ZodW_Metabat.232
ZodW_Metabat.241
Zod_Metabat.1203
Zod_Metabat.397
ZodW_Metabat.368
ZodW_maxbin.163
ZodW_Metabat.7
Zod_Metabat.65
Zod_Metabat.777
Zod_Metabat.800
ZodW_Metabat.117
Zod_Metabat.642
Zod_Metabat.1236
Zod_Metabat.217
Zod_Metabat.768
Zod_Metabat.943
Zod_Metabat.1005
Zod_Metabat.688
Zod_Metabat.37
Zod_Metabat.988
ZodW_Metabat.2
Zod_Metabat.810
Zod_Metabat.669
Zod_Metabat.1409
Zod_Metabat.64
ZodW_Metabat.13
Zod_Metabat.664
ZodW_Metabat.292
Zod_Metabat.612
Zod_Metabat.1255
ZodW_Metabat.155
Zod_maxbin.0112
ZodW_Metabat.353
Zod_Metabat.1419
ZodW_Metabat.19
Zod_Metabat.1376
Zod_Metabat.1198
ZodW_Metabat.59
ZodW_Metabat.94
Zod_Metabat.1042
ZodW_Metabat.248
ZodW_Metabat.183
ZodW_Metabat.166
Zod_Metabat.325
ZodW_Metabat.376
ZodW_Metabat.182
ZodW_Metabat.383
ZodW_maxbin.137
Zod_Metabat.545
ZodW_Metabat.217
ZodW_Metabat.319
Zod_Metabat.638
ZodW_Metabat.352
Tree scale: 0.1
Sedimentisphaerales Families
Anaerohalophaeraceae
Novel:ZNF071
SM23-30
C
Zod Metabat.1166
ZodW maxbin.277
Zod Metabat.1001
Zod Metabat.35
Zod Metabat.1436
Zod Metabat.1349
Zod Metabat.916
Zod Metabat.1380
Zod Metabat.331
Zod Metabat.769
Zod Metabat.1370
Zod Metabat.67
Zod Metabat.138
Zod Metabat.827
Zod Metabat.417
Zod Metabat.1071
Zod Metabat.289
Zod Metabat.999
Zod Metabat.1083
Zod Metabat.96
Zod Metabat.683
Zod Metabat.1411
Zod Metabat.1017
Zod Metabat.706
Zod Metabat.1213
Zod Metabat.1177
Zod Metabat.388
Tree scale: 0.1
Spirochaeales Families
ARS1246
Marispirochaetaceae
Novel:ZNF082
Novel:ZNF103
RPPD01
Novel:ZNF083
Novel:ZNF084
Spirochaetaceae_B
Novel:ZNF102
D
Zod_Metabat.156
ZodW_Metabat.38
Zod_Metabat.752
Zod_Metabat.700
ZodW_Metabat.93
ZodW_Metabat.41
Zod_Metabat.1463
Zod_Metabat.1170
Zod_Metabat.745
Zod_Metabat.1342
Zod_Metabat.1377
Zod_Metabat.281
Zod_Metabat.934
Zod_Metabat.353
ZodW_Metabat.242
Zod_Metabat.1464
Zod_Metabat.855
Zod_Metabat.749
Tree scale: 0.1
Syntrophales Families
UBA2210
UBA5619
Novel:ZNF036
Novel:ZNF035
Smithellaceae
UBA3084
MLS-D
Syntrophaceae
UBA2185
E
Zod_Metabat.153
Zod_Metabat.402
Zod_Metabat.215
Zod_Metabat.836
Zod_Metabat.309
Zod_Metabat.41
Zod_Metabat.324
Zod_Metabat.27
Zod_Metabat.781
ZodW_Metabat.303
Zod_maxbin.0017_4
Zod_Metabat.729
Zod_Metabat.88
Zod_Metabat.996
ZodW_maxbin.216
Zod_Metabat.1020
Zod_Metabat.1365
Zod_Metabat.717
Tree scale: 0.1
Woesearchaeales Families
B72-G16
Novel:ZNF08
Novel:ZNF107
UBA11576
Novel:ZNF106
CG08-08-20-14
GCA-2685855
Novel:ZNF010
UBA10107
Novel:ZNF09
SM23-78
GW2011-AR15
Novel:ZNF07
SKGA01
J091
F
ZodW_maxbin.125
Zod_Metabat.396
Zod_Metabat.513
Zod_Metabat.644
Zod_Metabat.161
Zod_Metabat.1039
Zod_Metabat.140
Zod_Metabat.1155
Zod_Metabat.825
Zod_Metabat.908
Zod_Metabat.833
Zod_Metabat.297
Zod_Metabat.390
Zod_Metabat.845
Zod_Metabat.265
Zod_Metabat.828
Zod_Metabat.712
Zod_Metabat.157
Zod_Metabat.1084
Zod_Metabat.822
Zod_Metabat.672
Tree scale: 0.1
Abundance in Gtdb
Abundant (>5 genomes)
Rare (5 genomes or less)
Novel (Absent in Gtdb)
Cultured versus Uncultured
Cultured Family
Uncultured Family


A
Sox
Sox
Dsr/Asr
H_2S Sulfide
Psr/ Hyd/Shy
Fcc/Sqr
Elemental sulfur/ Polysulfide
rDSR
SO_3^{2-} Sulfite
Sat+Apr
Apr/Sox/ Soe/Sor
SO_4^{2-} Sulfate
phs
rhd
Sox
rDSR
$S_2O_3^{2-}$ Thiosulfate
Otr/ttr
tsdA/ doxAD
$S_4O_6^{2-}$ Tetrathionate
tetH
Reduction
Oxidation
Disproportionation


B
% of genomes encoding the genes
Hyd/Shy
Psr
Apr
Dsr
Asr
rhd
otr
ttr
phs
Soe
Sor
tetH
Sox
rDSR
fccAB
SQR
tsdA
doxAD
S_2O_3 disproportionation
S_4O_6 reduction to sulfide
S_2O_3 reduction
Sulfide oxidation
S^0 oxidation
Sulfate reduction
S species oxidation
Sulfite oxidation
Polysulfide reduction
Sulfite reduction
S_4O_6 oxidation
S_2O_3 oxidation
Sediment
water
Acidobacteriota
Asgardarchaeota
Bdellovibrionota
Cloacimonadota
Eisenbacteria
Fusobacteriota
Latescibacterota
Nanoarchaeota
Planctomycetota
Sumerlaeota
UBP14
Zixibacteria
Actinobacteriota
Aureabacteria
Bipolaricaulota
CSSED10-310
FCPU426
Halobacteriota
Marinisomatota
Nitrospirota
Proteobacteria
Synergistota
UBP6
Aenigmatarchaeota
B130-G9
Calditrichota
Delongbacteria
Fermentibacterota
Armatimonadota
BMS3Abin14
Chloroflexota
Desulfobacterota
Firmicutes_A
KSB1
Myxococcota
Patescibacteria
Spirochaetota
Thermoproteota
WOR-3
Iainarchaeota
Margulisbacteria
OLB16
RBG-13-66-14
SZUA-182
UBP7_A
Altarchaeota
Bacteroidota
Campylobacterota
Dependentiae
Fibrobacterota
Krumholzibacteriota
Methanobacteriota
Omnitrophota
SM23-31
Thermoplasmatota
Verrucomicrobiota

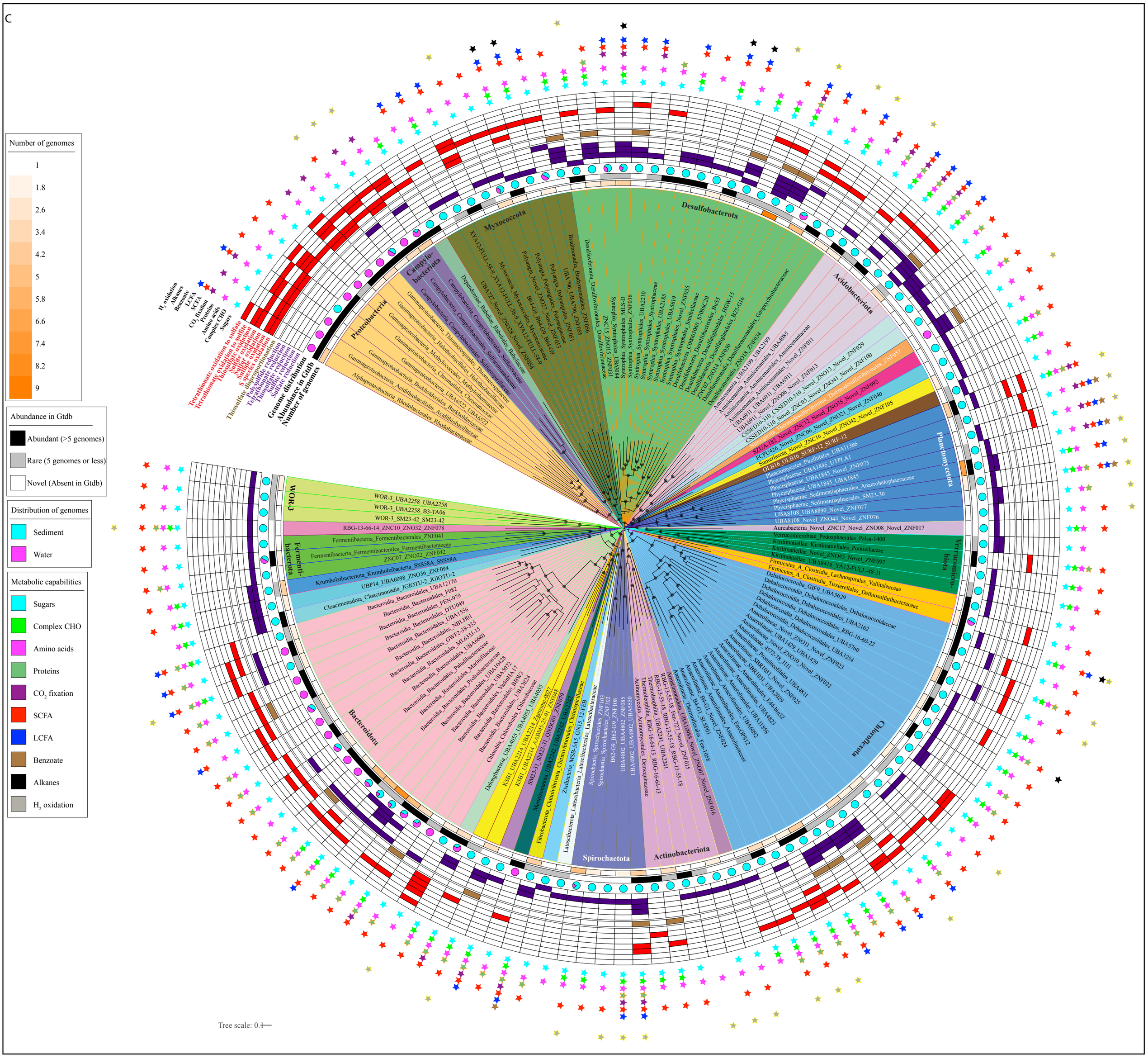
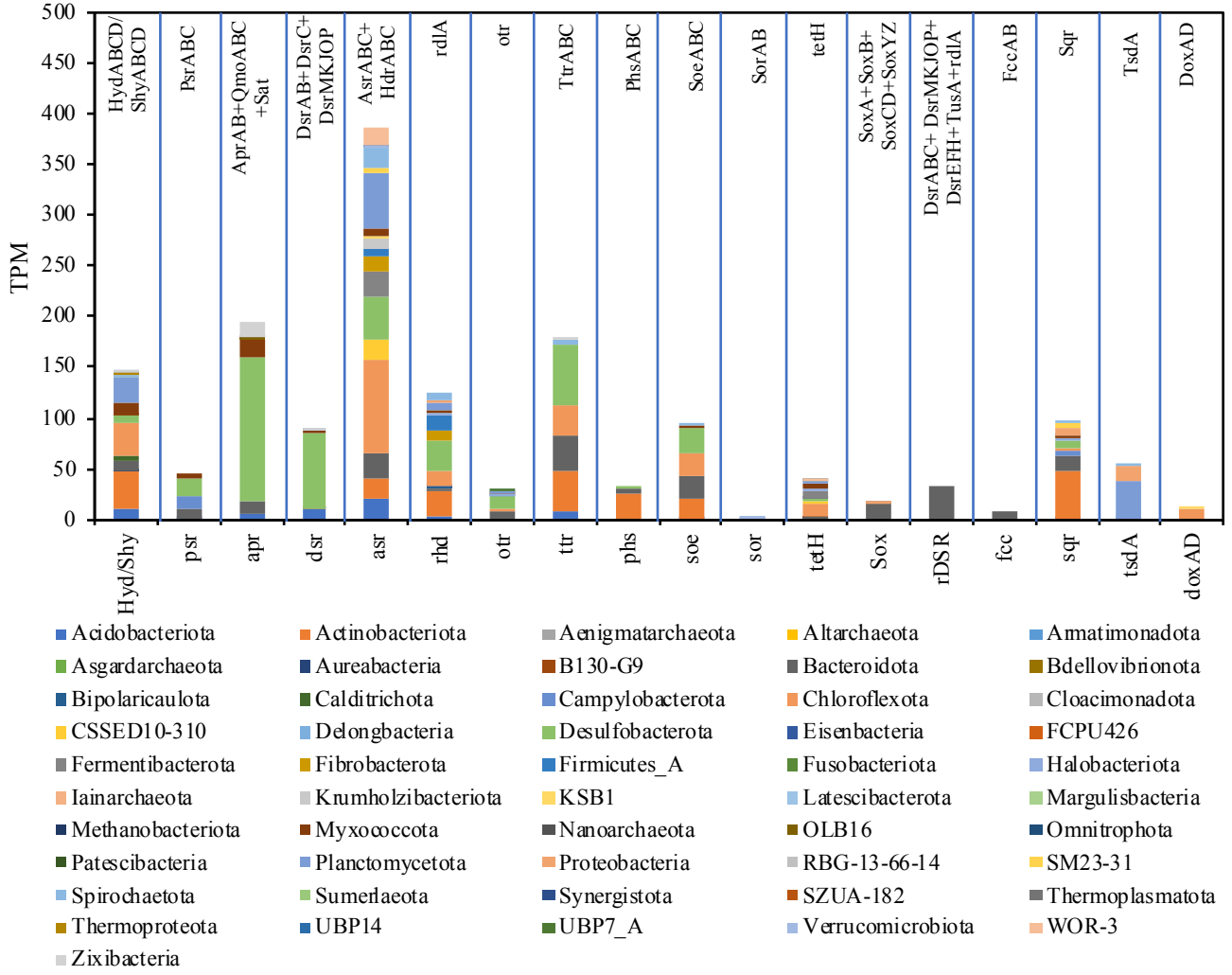

C
Number of genomes
1
1.8
2.6
3.4
4.2
5
5.8
6.6
7.4
8.2
9
Abundance in Gtdb
Abundant (>5 genomes)
Rare (5 genomes or less)
Novel (Absent in Gtdb)
Distribution of genomes
Sediment
Water
Metabolic capabilities
Sugars
Complex CHO
Amino acids
Proteins
CO_2 fixation
SCFA
LCFA
Benzoate
Alkanes
H_2 oxidation
H_2 oxidation
Alkanes
Benzoate
LCFA
SCFA
CO_2 fixation
Proteins
Amino acids
Complex CHO
Sugars
Tetrathionate oxidation to sulfate
Tetrathionate oxidation to sulfite
Thiosulfate oxidation
Sulfur oxidation
Sulfide oxidation
S species oxidation
Thiosulfate disproportionation
Polysulfide reduction
Tetrathionate reduction
Sulfate reduction
Thiosulfate reduction
Sulfite reduction
Genome distribution
Abundance in Gtdb
Number of genomes
Desulfobacterota
Acidobacteriota
Myxococcota
Campylo-bacteriota
Proteobacteria
Planctomycetota
Verrucomicro-biota
Chloroflexota
Actinobacteriota
Spirochaetota
Latescibacterota
Zixibacteria
Fibrobacterota
Marinisomatota
Bacteroidota
Fermenti-bacteriota
WOR-3
Tree scale: 0.1

Figure 6


TPM
500
450
400
350
300
250
200
150
100
50
0
HydABCD/ ShyABCD
PsrABC
AprAB+QmoABC +Sat
DsrAB+DsrC+ DsrMKJOP
AsrABC+ HdrABC
rdlA
otr
TtrABC
PhsABC
SoeABC
SorAB
tetH
SoxA+SoxB+ SoxCD+SoxYZ
DsrABC+ DsrMKJOP+ DsrEFH+TusA+rdlA
FccAB
Sqr
TsdA
DoxAD
Hyd/Shy
psr
apr
dsr
asr
rhd
otr
ttr
phs
soe
sor
tetH
Sox
rDSR
fcc
sqr
tsdA
doxAD
Acidobacteriota
Actinobacteriota
Aenigmatarchaeota
Altarchaeota
Armatimonadota
Asgardarchaeota
Aureabacteria
B130-G9
Bacteroidota
Bdellovibrionota
Bipolaricaulota
Calditrichota
Campylobacterota
Chloroflexota
Cloacimonadota
CSSED10-310
Delongbacteria
Desulfobacterota
Eisenbacteria
FCPU426
Fermentibacterota
Fibrobacterota
Firmicutes_A
Fusobacteriota
Halobacteriota
Iainarchaeota
Krumholzibacteriota
KSB1
Latescibacterota
Margulisbacteria
Methanobacteriota
Myxococcota
Nanoarchaeota
OLB16
Omnitrophota
Patescibacteria
Planctomycetota
Proteobacteria
RBG-13-66-14
SM23-31
Spirochaetota
Sumerlaeota
Synergistota
SZUA-182
Thermoplasmatota
Thermoproteota
UBP14
UBP7_A
Verrucomicrobiota
WOR-3
Zixibacteria