The evolution and spread of sulfur-cycling enzymes reflect volcanic sulfur sources and the redox state of the early Earth

Katherine Mateos*[1,2], Garrett Chappell*[1,3], Eva Stüeken[4], and Rika Anderson[1, 5†]

[1] *Carleton College, Northfield, MN, USA*
[2] *Ocean Sciences Department, University of California Santa Cruz, Santa Cruz, CA, USA*
[3] *Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA*
[4] *University of St. Andrews, School of Earth & Environmental Science, Bute Building, Queen's Terrace, St Andrews, Fife, KY16 9TS, United Kingdom*
[5] *NASA NExSS Virtual Planetary Laboratory, University of Washington, Seattle, WA, USA*

[†] *Corresponding author:*
Rika Anderson
Carleton College, One North College Street, Northfield, MN, 55057 USA
507.222.4382
randerson@carleton.edu

**These authors contributed equally to this manuscript.*

Classification: Physical Sciences: Earth, Atmospheric, and Planetary Sciences; Biological Sciences: Evolution, Microbiology

Keywords: sulfur, biogeochemical cycles, early Earth evolution, microbial metabolism

Abstract

The biogeochemical sulfur cycle plays a central role in fueling microbial metabolisms, regulating the redox state of the Earth, and impacting climate through remineralization of organic carbon. However, traditional reconstructions of the ancient sulfur cycle based on geochemistry are confounded by ambiguous isotopic signals, low sulfate concentrations in the Archean ocean, and the isotopic impacts of photolysis acting on volcanogenic SO_2 gas. Here, we use a phylogenomics approach to ascertain the timing of gene duplication, loss, and horizontal gene transfer events for sulfur cycling genes across the tree of life. Our results suggest that metabolisms using sulfate reduction and sulfide oxidation emerged early in life's evolution, but metabolic pathways involving thiosulfate and the *sox* pathway proliferated across the tree of life only after the Great Oxidation Event, suggesting enhanced recycling of sulfur with increasing oxygen levels in the Paleoproterozoic ocean. Moreover, our results provide the first indication of organic sulfur cycling in the Archean with implications for climate regulation and atmospheric biosignatures. Overall, our results provide new insights into how the biological sulfur cycle evolved in tandem with the redox state of the early Earth and illustrate the power of combining geochemical and bioinformatics approaches to understand the evolution of Earth over time.

Significance statement

Sulfur is a crucial element for life. It plays a central role in regulating Earth's climate and redox state, and it was likely involved in life's earliest metabolic pathways. Microbes have played a key role in mediating the global sulfur cycle since the origin of life itself, yet geochemical methods are limited in their ability to reconstruct biological sulfur metabolisms over time. By constraining the timing of when different genes in the biological sulfur cycle were born and then spread across the tree of life, we can learn when specific metabolic pathways arose and became ecologically important on the early Earth. Our results illustrate how the evolution of the biological sulfur cycle mirrored the redox state of the early Earth, and provide the first indication that organic sulfur cycling originated during the Archean. These results provide important insights into how microbial pathways co-evolved with the environment on the early Earth, with important implications for atmospheric biosignatures and Earth's climate.

Introduction

The biogeochemical sulfur cycle has played a crucial role in the evolution of life and surface processes over geologic time. Dissimilatory metabolisms, including elemental sulfur reduction, sulfate reduction, sulfate disproportionation, and sulfide oxidation fuel diverse microbes and play a significant role in regulating the redox state of the Earth (1) **(Figure 1**). For example, the burial of biogenic sulfide in marine sediments may have contributed to progressive oxygenation of surface environments (2). Conversely, the metabolic reduction of sulfate is often coupled with the oxidation of organic carbon, a process that accounts for up to 50% of organic carbon mineralization (3). Recent studies have found that the sulfur cycle is intricately entwined with cycling of other important elements, including nitrogen and various transition metals (4, 5). In particular, freely dissolved sulfide in seawater, especially during the Proterozoic eon (6), may have impacted the solubility of essential micronutrients such as molybdenum (7). Thus, a deeper understanding of the evolution of the biological sulfur cycle can offer important insights into the evolution of other biogeochemical cycles and the oxidation state of our planet over time.

Several geochemical studies suggest that sulfur metabolisms were probably among the earliest metabolisms on the ancient Earth. Early analyses of sulfur isotopes of pyrite and barite in the 3.5 Ga Dresser formation provided evidence for sulfate reduction in the Archaean era (8). While data from Philippot et al. (9) later argued that the isotopic fingerprint of the sulfides was more consistent with sulfur disproportionation, subsequent work supported the original claim (10).

Since then, several studies have documented isotopic evidence of microbial sulfur cycling in a variety of Archean environments (e.g. (11–13)). However, this approach has limitations for three reasons: First, the isotopic signatures of differing sulfur metabolisms are not necessarily distinct enough to be recognizable in the sedimentary rock record. For example, analysis from the 3.2 Ga Moodies Group in South Africa confirmed the presence of reductive sulfur cycling, and hinted at the presence of oxidative sulfur cycling (14); however, the latter could not be unambiguously inferred. Second, the concentration of sulfate in the Archean ocean was low, possibly as low as 2.5 uM (15), dampening the signal of microbial sulfur isotope fractionation in the rock record. This challenge was illustrated by analysis of sulfur isotope ratios in a modern sulfate-poor analog of the Archaean ocean, where biological fractionations are muted despite the presence of active microbial sulfate reduction (15). Third, the Archean sulfur isotope record is famously impacted by photochemical processes acting on volcanogenic SO_2 gas in an anoxic atmosphere (16). These photochemical reactions are recognizable by significant so-called mass-independent isotopic fractionations; however, distinguishing these from biogenic isotope effects requires analyses of all four stable isotopes of sulfur and relatively large sample sets (10, 17, 18). Thus, while the geochemical record has produced valuable insight into sulfur cycling in the Archean, the results can be inconclusive and often cannot distinguish between specific metabolic pathways.

Given the significant limitations and uncertainties presented by the geochemical record, pairing geochemical analysis with top-down phylogenetics approaches can provide a novel perspective into the emergence and spread of distinct sulfur-cycling microbial metabolisms on the early Earth. Some molecular work has been conducted to explore the history of microbial sulfur cycling, with a particular focus on dissimilatory sulfate reduction (19). The enzyme DsrAB is used to both reduce and oxidize sulfite to sulfide in sulfate-reducing bacteria. The reduction of sulfite by DsrAB is considered the rate-limiting step in the microbial sulfate reduction pathway. A phylogenetic analysis of DsrAB identified three major clades of the gene: the reductive bacterial type, the oxidative bacterial type, and the reductive archaeal type (19). The reductive archaeal-type

proteins were most deeply rooted within the DsrAB tree, suggesting that the reductive pathway predated the oxidative one. Furthermore, the oxidative bacterial type DsrAB proteins was found to form a monophyletic group, suggesting that this protein evolved from a reductive-type ancestor. These data are consistent with low atmospheric oxygen concentrations on the Archean Earth, where a reductive pathway would be favored, and they are also in line with geochemical evidence for the early emergence of sulfur/sulfate-reducing bacteria. However, the authors of that study did not attempt to constrain the timing of these evolutionary events, making it difficult to draw concrete conclusions about how these proteins evolved in the context of biogeochemical transitions on the early Earth.

One of the biggest questions regarding the evolution of the biological sulfur cycle is how it co-evolved with the oxygenation of the Earth over time. It is now well established that the Earth's surface underwent a major transformation at around 2.4 Ga, when atmospheric O_2 levels increased above a threshold of 10^{-5} times present levels, known as the Great Oxidation Event (GOE) (20). As a consequence, volcanism was replaced by oxidative weathering as the major source of sulfur to the ocean, and the speciation of this new sulfur source was dominated by sulfate, as opposed to volcanogenic SO_2, dissolved sulfite, and the photochemical products S_8 and sulfate (21). Further, O_2 became abundant in surface waters as a potent oxidant of reduced sulfur species (22). The Archean-Proterozoic transition also witnessed a decline in hydrothermal activity on the ocean floor (23). Lastly, the deep ocean became fully oxygenated in the Neoproterozoic or early Paleozoic with the second rise of oxygen, leading to enhanced sulfide oxidation within sediments (24). Geochemical data show isotopic expressions of these events in the sulfur cycle (1); however, it is so far unknown if these isotopic signals reflect merely an enhancement of a pre-existing process or true evolutionary innovations. This question has important implications for cause-effect relationships in Earth system evolution.

Previous studies have used phylogenetic approaches to track the birth of specific lineages or genes (e.g. (25, 26); however, the birth of a gene does not necessarily coincide with the time at which the function of that gene became ecologically important. As genes are horizontally transferred between divergent microbial lineages, the genes that serve a useful function are the ones most likely to be retained in a genome (27–32). Thus, an increase in horizontal gene transfer events for a particular gene at a specific point in time likely gives an indication the gene in question was ecologically important during that time period.

To examine the evolution of the biological sulfur cycle over time, we therefore used phylogenomics approaches to track the timing of duplication, loss, and horizontal gene transfer events for sulfur cycling genes across a time-calibrated tree of life. This analysis allows us to determine approximately when these genes first arose and then proliferated across the tree of life on the early Earth. A similar analysis of nitrogen-cycling genes revealed that nitrogen fixation arose and spread early, while genes related to denitrification from nitrite arose and spread much later in Earth history (33). Here, we focus on constraining the timing of duplication, loss, and horizontal gene transfer events for genes related to dissimilatory sulfate reduction and sulfide oxidation via sulfide, transformations between sulfate and thiosulfate, as well as organic sulfur cycling.

Methods

Genome selection and construction of species tree

To construct the species tree, we included one representative genome from each bacterial and archaeal order, based on GTDB taxonomy (34, 35). Some eukaryotic genomes were also included to ensure a robust tree

topology, but the focus of the study was on sulfur cycling genes within bacterial and archaeal genomes. GToTree (36) was applied to identify and align single-copy universal ribosomal genes from the genomes we selected. The concatenated gene alignments were created from a set of fifteen universal single-copy genes (37), and we excluded genomes with fewer than half of the single-copy genes. Briefly, the GToTree workflow used prodigal (38) to predict genes on input genomes, then identified genes with HMMER3 v3.2.2 (39), individually aligned genes with MUSCLE v5.1 (40), trimmed the alignment with trimal v1.4.rev15 (41), and concatenated aligned genes with FastTree2 v2.1.1 (42). The resulting alignment was used to construct a phylogeny using RAxML v. 8.2.9 (43) with 100 rapid bootstraps using the PROTGAMMALG model of evolution as per (37). The root of the tree was placed in the bacterial domain (44). The resulting tree contains 871 genomes, including 777 bacterial, 80 archaeal, and 14 eukaryotic genomes.

Construction of time-calibrated chronogram

The species tree was converted to a chronogram using Phylobayes (45). We tested two separate sets of calibration points, one conservative (which represents the earliest date for which there is the most consensus for a given event based on the current scientific literature) and one liberal (which represents the earliest date for which there is any evidence of a given event based on the current scientific literature) to test the sensitivity of methodology (**Table 1**). The root age was set via a normally distributed gamma root prior according to calibration points in Table 1 with a standard deviation set to 200 million years, consistent with previous studies (26).

To generate chronograms, we tested three different clock models: autocorrelated log normal (LN) (46), uncorrelated gamma multiplier (UGAM) (47), and the autocorrelated CIR process (48). For each model and set of calibration points, two chains were run concurrently and were compared as a test of convergence. We analyzed convergence using the tracecomp program in Phylobayes, requiring an effective size >100 and a maximum difference between chains of <0.3. Each chain was run for >60,000 cycles. Chronograms were generated using the readdiv function in Phylobayes, discarding the first 2500 cycles as a burn in.

Identification of sulfur cycling genes and construction of gene trees

We identified sulfur cycling genes of interest using the sulfur metabolism pathway on the Kyoto Encyclopedia of Genes and Genomes (KEGG) (49–51). In line with previous genomic studies of the sulfur cycle, we analyzed dissimilatory sulfate reduction/sulfide oxidation genes (*aprAB*, *dsrAB*) (Anantharaman et al. 2018) and thiosulfate oxidizing/sulfate reducing genes (*soxABCXYZ*) (Canfield et al. 2010). Though the *sat* gene catalyzes the first step of dissimilatory sulfur cycling, we excluded this gene from our study because it is also used in other metabolic pathways that were not of specific interest here. We also examined genes that were involved in the production of volatile organic sulfur compounds, including methanethiol, dimethyl sulfoxide and dimethyl sulfide (*mddA, dmdA, dmsA*). The list of sulfur-cycling genes analyzed here was not meant to be exhaustive, but rather focused on core genes involved in dissimilatory sulfur oxidation and reduction for energy acquisition as well as select genes involved in organic sulfur cycling.

For consistency in identifying genes across genomes, we used Annotree (52) to identify sulfur cycling genes in microbial genomes using KEGG orthology numbers as queries (**Table 2**). We limited our analysis to the core dissimilatory sulfur cycling and thiosulfate reduction and oxidation genes that were included in the Annotree database. It is important to note that some metabolic pathways use the same genes for catalyzing oxidation and reduction reactions, and thus cannot be distinguished using these methods: for example, the *sox* operon is involved in both thiosulfate oxidation and sulfur disproportionation. The default Annotree settings

(minimum 30% identity, maximum e-value of 10^{-5}, minimum 70% subject and query alignment) were applied for identifying genes in genomes. Annotree output was curated to only include genes from genomes within our species tree. Note that although eukaryotic genomes were included in the species tree to ensure accurate topology, AnnoTree does not include eukaryotic phyla in the gene distribution search, and so eukaryotic genes were excluded from this analysis. The number of hits for each gene can be found in **Table 2**. These genes were aligned using MUSCLE v3.8.31 (53) then trimmed using TrimAl v.1.3 (41) with the -automated1 option as implemented in Phylemon2 (54). All gene phylogenies were constructed using RAxML-NG v. 0.9.0 (55). The model of evolution was determined by the Model Selection tool implemented in IQ-TREE 2 (56) with default settings. Trees were run with at least 1000 bootstraps or until the MRE-based bootstrapping test implemented by RAxML-NG fell below a cutoff of 0.03. The number of bootstraps used for each gene tree is reported in **Table 2**. In some cases, we used fewer bootstraps than required to reach convergence when performing reconciliation with ecceTERA due to computational limitations. These are noted in **Table 2**.

Reconciliation of gene trees with species chronogram with ecceTERA

Gene trees and species trees were reconciled using ecceTERA v1.2.5 (57) to identify gene loss, duplication, and transfer events. We used the default settings implemented in ecceTERA and amalgamated the gene trees (amalgamate=true). The output was configured to recPhyloXML format (58) with the option "recPhyloXML.reconciliation=true". Reconciliation analyses were performed on fully dated species trees and full sets of gene tree bootstraps where computational limits allowed (see **Table 3** for bootstrap information). Using a combination of custom Python scripts developed for this project (provided on GitHub at https://github.com/carleton-spacehogs/sulfur) and scripts provided in Duchemin et al. 2018, we calculated the mean date for each event based on the midpoints of the two 95% confidence intervals that defined the nodes of the branch on which the event occurred. Distributions of the gene event data produced from ecceTERA were subsequently compared to distribution data of the gene event data produced from AnGST (59) (see Supplemental Materials) to ensure the results were not dependent on the reconciliation algorithm.

It is important to consider the limitations inherent in dating each of these gene events. The phylogenetic reconciliation identifies branches on which events occurred, and thus there is no way to determine when on a given branch a specific gene event occurred. Combined with the inherent error associated with dating events on chronograms dating back billions of years, estimates of when gene events occurred should not be taken as absolute dates. Instead, we emphasize the relative timing of these events. By examining the distributions of when specific gene events occurred, we are able to better understand the relative timing of when specific metabolisms became ecologically significant.

Results

Construction of species tree and time-calibrated chronogram

We constructed a species tree from an alignment of fifteen universal single-copy ribosomal genes in order to conduct the phylogenetic reconciliation (**Figure 2**). The resulting tree placed the Eukaryotes within the Archaeal domain, consistent with a two-domain tree of life, as has been recovered previously using similar methods (60–62). We constructed chronograms from this species tree using two autocorrelated clock models (LN and CIR) and one uncorrelated clock model (UGAM). We tested both liberal and conservative fossil calibration points to construct the molecular clock (**Table 1**; see Methods). The liberal calibration points returned unrealistic ages for the last universal common ancestor (LUCA), and so were not used for the remainder of this analysis. Using the conservative calibration points, the LN clock returned a LUCA age of

approximately 4.48 Ga, the CIR clock returned a LUCA age of approximately 4.05 Ga; and the UGAM clock returned a LUCA age of approximately 3.93 Ga. For the remainder of our analysis we report the results from the CIR clock in the main text because the autocorrelated CIR results were deemed more realistic than the autocorrelated LN results, and autocorrelated clock models have previously been shown to outperform uncorrelated models such as UGAM (48); but all results from the LN and UGAM clocks are reported in the Supplementary Materials. All Newick files, alignments, and chronograms with error bars have been deposited in FigShare at https://figshare.com/account/home#/projects/144267.

Phylogenetic distribution of sulfur-cycling genes

We used AnnoTree (52) to determine the distribution of sulfur-cycling genes across the tree of life. Genes related to dissimilatory sulfate reduction and sulfide oxidation via sulfite, including *dsr* and *apr* genes, tended to be fairly widespread across the tree of life: *aprAB* in particular is fairly widespread, occurring in approximately 47 bacterial and 5-6 archaeal phyla, and *dsrAB* is found in 32 bacterial and 4-5 archaeal phyla. In contrast, genes related to thiosulfate oxidation/reduction, particularly the *sox* group of genes, were more phylogenetically restricted, occurring in approximately 14-20 bacterial and 1 archaeal phylum, with the majority of gene hits restricted to the Proteobacteria superphylum. The exceptions to this rule were *soxB* and *soxC*, which were much more widespread across the tree of life, occurring in approximately 31-38 bacterial and 4 archaeal phyla. Finally, the organic sulfur cycling genes *dmdA*, *dmsA*, and *mddA* each displayed a different phylogenetic distribution: *dmdA* was found in only 13 bacterial and 2 archaeal phyla, restricted mostly to the Proteobacteria and Actinobacteria; *dmsA* was much more widespread, identified in 43 bacterial and 6 archaeal phyla, but generally not observed in the Patescibacteria; and *mddA* was similarly widespread, found in 34 bacterial and 4 archaeal phyla, most noticeably absent from the Patescibacteria and the Firmicutes phyla.

Identification of duplication, loss, and horizontal gene transfer events for sulfur-cycling genes

For the 13 genes of interest, we quantified gene duplication, loss and horizontal gene transfer events using multiple reconciliation algorithms. Reconciliation was performed by comparing the topology of the maximum likelihood gene trees for each gene to fossil-calibrated chronograms using three different clock models (CIR, UGAM, and LN) using the reconciliation programs AnGST (59) and ecceTERA (57). The analyses presented below focus on results from ecceTERA based on the CIR clock model (see **Table 2**), with replicate analyses with similar results from ecceTERA using the UGAM and LN clock models as well as results from AnGST using the CIR, UGAM, and LN models presented in **Supplementary Tables 1-3 and Supplementary Figures 1-3**.

According to the earliest events identified for each of the genes, reported as the approximate gene birth date in **Table 3**, the most ancient genes include those involved in dissimilatory sulfate reduction and oxidation: *dsrAB,* which catalyzes the oxidation and reduction of sulfite and sulfide, and *aprAB*, which catalyzes the oxidation and reduction of adenylyl sulfate (APS) and sulfite. Arising slightly later are the *sox* genes (excluding soxX and soxY), which are involved in sulfate and thiosulfate ($S_2O_3^{2-}$) oxidation. The gene *mddA,* involved in the conversion of methanethiol (CH_3SH) to dimethylsulfide ($(CH_3)_2S$), arose slightly later. (See **Figure 1** for a schematic of the sulfur cycle and the steps catalyzed by each of these genes.) Other than *soxABXZ*, in which the majority of the events are losses, the gene events are largely dominated by horizontal gene transfers, with very few gene duplications across all the genes. For the *sox* genes, aside from *soxC*, there appears to be a small peak in gene transfer/loss/duplication events around 1.5 Ga (Figure 3). *soxC* is different from the rest of the *sox* genes, having an earlier birth and proliferation while also lacking the early peak. Additionally, *soxC* had

many more gene hits and gene events than the rest of the *sox* genes, and was found in a wider range of organisms overall.

The organic sulfur cycling genes *dmdA* for the conversion of dimethyl sulfide ($(CH_3)_2S$, DMS) into methanethiol (CH_3SH) and *dmsA* for conversion between DMS and dimethyl sulfoxide ($(CH_3)_2SO$, DMSO) appeared to be much younger than the genes involved in sulfur oxidation and reduction for energy metabolism, appearing only at approximately 1.7-1.2 Ga (**Figure 3**). In contrast, *mddA*, involved in the conversion of methanethiol to DMS, appeared much earlier according to our analyses, at almost 3 Ga.

Discussion

Our analysis of gene duplication, transfer, and loss events for sulfur cycling genes across Earth history provides insight into the relative timing for the proliferation of these genes across the tree of life, and thus has implications for when specific sulfur metabolisms became ecologically important. As has been suggested previously, if a gene is acquired via horizontal gene transfer and retained in the genome, this indicates that the horizontally transferred gene likely has been selected and retained because it performs a useful ecological function (27–32, 75). Thus, a rise in horizontal gene transfer events for a specific gene at a given time can indicate when these genes became favorable or ecologically useful given the conditions of the environment at that point in Earth's history (33). A similar analysis presented in (33) revealed a gradual shift in the biological nitrogen cycle over Earth history.

It is important to note that the patterns across all of the genes presented here show a peak of gene events around 750 Ma. However, many of these events occurred on long branches terminating in leaves of the species chronogram, meaning that these specific events could have occurred anywhere on that branch, up to present day. Due to limitations inherent in species and gene tree reconciliation, the date of each event can only be confined to the branch of the species tree where it occurred, meaning that it could have happened at any point in time between the two nodes of that branch. The data reported for events occurring on leaves reflect the midpoints between the terminus of the leaf and the last node of the leaf, many of which occurred at approximately 1.5 Ga. Thus, although many events were calculated to have occurred around 750 million years ago, in reality these events occurred at some point during a prolonged time period that extends to the present day. Due to these limitations in the data, caution is warranted in interpreting these relatively more recent gene events, as the calculated dates are rough estimates by necessity. Thus, our analysis of the histograms of gene events will emphasize relative trends in the frequency of events over time, rather than focusing on precisely dated events. Additionally, the approximate dates for gene birth events are inferred based on the earliest event for that gene in the reconciliation, so the birth of that gene occurs prior to the earliest event by definition.

Dissimilatory sulfur reduction and oxidation

Metabolisms involving dissimilatory sulfate reduction and sulfide oxidation involve the genes *dsrAB* and *aprAB*, with adenylyl sulfate (APS) and sulfite as intermediates. Our results indicate that these genes are more ancient than the *sox* genes and some of the organic cycling genes, beginning to proliferate at around 3 Ga (**Figure 3**). These results are consistent with geochemical studies suggesting that dissimilatory sulfate reduction is ancient (8). At that time, the major source of sulfate would have been photochemical oxidation of volcanic SO_2 gas (12, 16), and dissolution of SO_2 in water would have generated sulfite (21). The reduction

of both sulfate and sulfite could have been coupled to organic matter oxidation, as well as to volcanogenic or biogenic H_2 oxidation. Reduced forms of sulfur would have been abundant prior to the Great Oxidation Event, particularly near hydrothermal vent systems (76), and their biological oxidation may have been coupled to the reduction of Fe^{3+} or trace O_2 that occurred locally in the surface ocean. Furthermore, phototrophic sulfide oxidizers use some of the same genes for sulfide oxidation as chemotrophs, including *sox* and *dsr* genes (77, 78). Thus our data allow for the possibility that significant sulfide oxidation was carried out at the sea surface by phototrophic sulfide oxidizers, not requiring other sources of oxidants. The findings that the *dsrAB* genes have an ancient origin are consistent with phylogenetic results from Wagner et al. 1998, who used targeted gene sequencing and 16S rRNA sequencing techniques to show that dissimilatory sulfite reductase genes originated at around 3 Ga (79). Although *aprAB* does not involve sulfide directly, the production of sulfite from sulfide using the *dsrAB* genes may have prompted the production of APS using *aprAB*. Previous researchers have theorized that both the *apr* and *dsr* genes were involved in early oxidative pathways using sulfide in ancient microbial mats around 3 billion years ago (80).

Sulfate-thiosulfate transformations

The Sox enzyme system is involved in the reduction and oxidation of sulfate and thiosulfate, respectively (81). A version of this pathway, omitting the SoxCD complex, can also be used to oxidize hydrogen sulfide to elemental sulfur (1). Our results indicate that, like the other genes included in our analysis, the *sox* genes have a peak in the number of events at around 750 Ma. While this may be a bioinformatics artifact, it also coincides with the Neoproterozoic Oxygenation Event (NOE, (20)), where the deep ocean is thought to have become more pervasively oxygenated. However, most *sox* genes (except for *soxC*) also have an earlier peak at around 1.5 Ga. While it is unclear why this is the case, it may be related to the fact that *soxC* is not involved in the alternate *sox* pathway that creates elemental sulfur from sulfide. The *soxC* gene is part of a sulfur dehydrogenase molybdenum enzyme complex called *soxCD* that catalyzes a six-electron transfer in the middle of the *sox* sequence, and appears to be reliant on the other enzyme complexes in the *sox* sequence (78, 82). However, while *soxC* had a different pattern of gene duplication/loss/transfer events through its evolutionary history compared to the other *sox* genes, it also had considerably more gene hits and gene events than the other *sox* genes analyzed. *soxC* exists in a wider range of organisms than the rest of the *sox* genes, which were primarily found in Proteobacteria. We speculate that this pattern may be the result of the gene's relationship to another gene with a similar function, *sorA*, which has a 26.5% sequence identity to *soxC* (82) and is similarly widespread across the tree of life. Alternatively, it could indicate a separate function beyond the *sox* pathway for *soxC* that would require further investigation.

Setting aside the complexities around *soxC*, the genes *dsr* and *apr* appear to have arisen earlier than genes involved in the *sox* complex. The number of gene duplication, loss, and horizontal gene transfer events for *soxABXYZ* experienced a sharp increase around 2 billion years ago, which approximately coincides with increasing sulfate availability in the Earth's oceans after the GOE (83). Thiosulfate has an intermediate redox state (S(+II)) between sulfide (S(-II)) and sulfate (S(+VI)) and forms most commonly during microbial sulfide oxidation (84). Hence the expansion of *sox* genes across the tree of life in the Paleoproterozoic is most parsimoniously attributed to increasing availability of O_2 and therefore enhanced sulfide oxidation. This finding is consistent with geochemical evidence for enhanced disproportionation of elemental sulfur in the mid- to late-Proterozoic (22, 24), as elemental sulfur, like thiosulfate, is an intermediate in microbial sulfide oxidation. Our data thus indicate that these intermediates became more abundant and thus more favorable metabolic substrates from the Paleoproterozoic onwards. In other words, these results indicate that the GOE and NOE indirectly triggered metabolic innovations in the sulfur cycle.

The slight delay between the geochemically defined GOE at 2.4 Ga (85) and the peak in the expansion of *sox* genes could be the result of either delayed oxygen delivery to the deep ocean after the GOE, or a delay in the reflection of ecological changes in the gene record. A similar pattern has been observed in the diversification of cyanobacteria, at around 2 Ga, as a delayed reaction to the GOE (86). After this first peak, the number of gene events decreased over time for a few hundred million years, potentially indicating that this gene stabilized in a specific set of ecological niches. Today, these genes are mostly found in Proteobacteria across a range of ecological niches such as deep-sea hydrothermal vents, intertidal flats and marshes, soils, and brackish lagoons (87).

Organic sulfur cycling

The organic sulfur cycle involves the biological formation of volatile organic compounds such as dimethyl sulfide (DMS) and methanethiol. Both *dmdA* and *dmsA*, the key enzymes involved in DMS metabolisms, appear to be much younger than *mddA*, which converts methanethiol into DMS. The genes *dmdA* and *dmsA* record their first events at around 1.5 Ga, possibly linked to the rise of eukaryotic algae, whose production of organic sulfur gasses has been implicated in global cooling in the late Proterozoic (88). In contrast, *mddA* had already appeared at around 3 Ga, according to our analysis. Thus, our results suggest that bacteria were capable of generating DMS fairly early in Earth history, possibly with important implications for climate regulation on the early Earth, because DMS particles are known to act as cloud condensation nuclei with the effect of cooling the Earth's surface (89). Furthermore, the formation of DMS in the Archean may have had important implications for the generation of mass independent fractionation (MIF) of sulfur isotopes, which is abundant in the Archean rock record and thought to reflect photochemical reactions in the anoxic Archean atmosphere. Volcanogenic SO_2 is generally thought to have been the most important sulfur gas in the Archean (90); however, the potential role of organic gasses has been considered in theoretical studies (91, 92). Our results may thus be the first tentative evidence that such organic sulfur gasses were indeed produced on the Archean Earth and may therefore be important to consider in the interpretation of MIF records.

Modern sources of naturally occurring methanethiol (the substrate for DMS formation) include decaying organic matter and the breakdown of the algal metabolite DMSP (dimethylsulfonioproprionate, $C_5H_{10}O_2S$). On the early Earth, hydrothermal vents may have been an important source of methanethiol (93). Interestingly, thiolated organic compounds from hydrothermal vents have been invoked as key substrates in the evolution of metabolic pathways (specifically the acetyl-CoA pathway) leading to the origin of life (94). Theoretical calculations suggest that millimolar concentrations of methanethiol may be generated abiotically from H_2, CO_2 and H_2S in hydrothermal settings (95). The antiquity of the *mdd* gene may thus be indirect evidence for the importance of hydrothermal substrates in the origin and early evolution of the biosphere. Moreover, volatile organic sulfur compounds such as DMS are important as potential remotely detectable biosignatures, because they can conceivably be detected on other planets with an anoxic biosphere using analysis of the spectral signatures of the planet's atmosphere (96). Our results thus suggest that Earth's earliest biosphere may potentially have been detectable through this technique.

Conclusions

Shifts in the biological sulfur cycle have a profound impact on the global carbon cycle and Earth's climate and are closely tied to the redox state of the Earth. Our results confirm previous geochemical results suggesting that microbial energy acquisition via sulfate reduction and possibly sulfide oxidation emerged early in Earth

history, and they indicate that metabolisms involving intermediates such as thiosulfate proliferated across the tree of life only after the Paleoproterozoic Great Oxidation Event, as the Earth's ocean and atmosphere became more oxidizing. However, our analysis goes beyond the geochemical records, because our data reveal that the expressions of these geochemical signatures were not merely the result of preservation or expansion of a single organism but instead caused by the radiation of genomic innovations across the tree of life. Furthermore, our analysis provides the first indication for organic sulfur cycling in the Archean, possibly capitalizing on methanethiol generated in hydrothermal vent environments. The formation of DMS from methanethiol in the Archean may have had important implications for global climate as well as for the generation of the sulfur MIF signal during photolysis in the atmosphere.

Acknowledgements
We thank Celine Scornavacca for assistance with ecceTERA, and Karthik Anantharaman for advice regarding the biological sulfur cycle. KM was supported by the Dean of the College Office at Carleton College. Funding for RA and GC was provided by the Virtual Planetary Laboratory, supported by the NASA Astrobiology Program under grant 80NSSC18K0829 as part of the Nexus for Exoplanet System Science (NExSS) research coordination network. EES acknowledges funding from a NERC Frontiers grant (NE/V010824/1).

References

1. D. A. Fike, A. S. Bradley, C. V. Rose, Rethinking the Ancient Sulfur Cycle. *Annual Review of Earth and Planetary Sciences* **43**, 593–622 (2015).

2. A. J. Krause, *et al.*, Stepwise oxygenation of the Paleozoic atmosphere. *Nat. Commun.* **9**, 4081 (2018).

3. B. B. Jørgensen, Mineralization of organic matter in the sea bed—the role of sulphate reduction. *Nature* **296**, 643–645 (1982).

4. D. E. Canfield, *et al.*, A cryptic sulfur cycle in oxygen-minimum-zone waters off the Chilean coast. *Science* **330**, 1375–1378 (2010).

5. V. Boyko, *et al.*, Biogeochemical cycling of sulfur, manganese and iron in ferruginous limnic analog of Archean ocean. *Geochimica et Cosmochimica Acta* **296**, 56–74 (2021).

6. E. A. Sperling, *et al.*, Statistical analysis of iron geochemical data suggests limited late Proterozoic oxygenation. *Nature* **523**, 451–454 (2015).

7. C. T. Reinhard, *et al.*, Proterozoic ocean redox and biogeochemical stasis. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 5357–5362 (2013).

8. Y. Shen, R. Buick, D. E. Canfield, Isotopic evidence for microbial sulphate reduction in the early Archaean era. *Nature* **410**, 77–81 (2001).

9. P. Philippot, *et al.*, Early Archaean Microorganisms Preferred Elemental Sulfur, Not Sulfate. *Science* **317**, 1534–1537 (2007).

10. Y. Ueno, S. Ono, D. Rumble, S. Maruyama, Quadruple sulfur isotope analysis of ca. 3.5Ga Dresser Formation: New evidence for microbial sulfate reduction in the early Archean. *Geochimica et Cosmochimica Acta* **72**, 5675–5691 (2008).

11. I. Zhelezinskaia, A. J. Kaufman, J. Farquhar, J. Cliff, Large sulfur isotope fractionations associated with Neoarchean microbial sulfate reduction. *Science* **346**, 742–744 (2014).

12. D. L. Roerdink, P. R. D. Mason, J. Farquhar, T. Reimer, Multiple sulfur isotopes in Paleoarchean barites identify an important role for microbial sulfate reduction in the early marine environment. *Earth and Planetary Science Letters* **331-332**, 177–186 (2012).

13. N. McLoughlin, E. G. Grosch, M. R. Kilburn, D. Wacey, Sulfur isotope evidence for a Paleoarchean subseafloor biosphere, Barberton, South Africa. *Geology* **40**, 1031–1034 (2012).

14. S. Nabhan, J. Marin-Carbonne, P. R. D. Mason, C. Heubeck, In situ S-isotope compositions of sulfate and sulfide from the 3.2 Ga Moodies Group, South Africa: A record of oxidative sulfur cycling. *Geobiology* **18**, 426–444 (2020).

15. S. A. Crowe, *et al.*, Sulfate was a trace constituent of Archean seawater. *Science* **346**, 735–739 (2014).

16. J. Farquhar, H. Bao, M. Thiemens, Atmospheric Influence of Earth's Earliest Sulfur Cycle. *Science* **289**, 756–758 (2000).

17. A. L. Zerkle, M. W. Claire, S. D. Domagal-Goldman, J. Farquhar, S. W. Poulton, A bistable organic-rich atmosphere on the Neoarchaean Earth. *Nature Geoscience* **5**, 359–363 (2012).

18. J. Farquhar, *et al.*, Pathways for Neoarchean pyrite formation constrained by mass-independent sulfur isotopes. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 17638–17643 (2013).

19. A. L. Müller, K. U. Kjeldsen, T. Rattei, M. Pester, A. Loy, Phylogenetic and environmental diversity of DsrAB-type dissimilatory (bi)sulfite reductases. *ISME J.* **9**, 1152–1165 (2015).

20. T. W. Lyons, C. T. Reinhard, N. J. Planavsky, The rise of oxygen in Earth's early ocean and atmosphere. *Nature* **506**, 307–315 (2014).

21. M. W. Claire, *et al.*, Modeling the signature of sulfur mass-independent fractionation produced in the Archean atmosphere. *Geochimica et Cosmochimica Acta* **141**, 365–380 (2014).

22. D. T. Johnston, *et al.*, Active microbial sulfur disproportionation in the Mesoproterozoic. *Science* **310**, 1477–1479 (2005).

23. S. Viehmann, M. Bau, J. Elis Hoffmann, C. Münker, Geochemistry of the Krivoy Rog Banded Iron Formation, Ukraine, and the impact of peak episodes of increased global magmatic activity on the trace element composition of Precambrian seawater. *Precambrian Research* **270**, 165–180 (2015).

24. M. Kunzmann, *et al.*, Bacterial sulfur disproportionation constrains timing of Neoproterozoic oxygenation. *Geology* **45**, 207–210 (2017).

25. E. S. Boyd, *et al.*, A late methanogen origin for molybdenum-dependent nitrogenase. *Geobiology* **9**, 221–232 (2011).

26. C. Magnabosco, K. R. Moore, J. M. Wolfe, G. P. Fournier, Dating phototrophic microbial lineages with reticulate gene histories. *Geobiology* **16**, 179–189 (2018).

27. M. L. Coleman, S. W. Chisholm, Ecosystem-specific selection pressures revealed through comparative population genomics. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 18634–18639 (2010).

28. J. O. McInerney, A. McNally, M. J. O'Connell, Why prokaryotes have pangenomes. *Nat Microbiol* **2**, 17040 (2017).

29. A. Moulana, R. E. Anderson, C. S. Fortunato, J. A. Huber, Selection Is a Significant Driver of Gene Gain and Loss in the Pangenome of the Bacterial Genus Sulfurovum in Geographically Distinct Deep-Sea Hydrothermal Vents. *mSystems* **5** (2020).

30. M. F. Polz, E. J. Alm, W. P. Hanage, Horizontal gene transfer and the evolution of bacterial and archaeal population structure. *Trends Genet.* **29**, 170–175 (2013).

31. O. Popa, E. Hazkani-Covo, G. Landan, W. Martin, T. Dagan, Directed networks reveal genomic barriers and DNA repair bypasses to lateral gene transfer among prokaryotes. *Genome Res.* **21**, 599–609 (2011).

32. V. Daubin, Bacterial Genomes as New Gene Homes: The Genealogy of ORFans in E. coli. *Genome Research* **14**, 1036–1042 (2004).

33. C. Parsons, E. E. Stüeken, C. J. Rosen, K. Mateos, R. E. Anderson, Radiation of nitrogen-metabolizing enzymes across the tree of life tracks environmental transitions in Earth history. *Geobiology* **19**, 18–34 (2021).

34. D. H. Parks, *et al.*, A complete domain-to-species taxonomy for Bacteria and Archaea. *Nat. Biotechnol.* **38**, 1079–1086 (2020).

35. D. H. Parks, *et al.*, A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* **36**, 996–1004 (2018).

36. M. D. Lee, GToTree: a user-friendly workflow for phylogenomics. *Bioinformatics* **35**, 4162–4164 (2019).

37. L. A. Hug, *et al.*, A new view of the tree of life. *Nature Microbiology* **1** (2016).

38. D. Hyatt, *et al.*, Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).

39. S. R. Eddy, Accelerated Profile HMM Searches. *PLoS Comput. Biol.* **7**, e1002195 (2011).

40. R. C. Edgar, MUSCLE v5 enables improved estimates of phylogenetic tree confidence by ensemble bootstrapping https:/doi.org/10.1101/2021.06.20.449169.

41. S. Capella-Gutiérrez, J. M. Silla-Martínez, T. Gabaldón, trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).

42. M. N. Price, P. S. Dehal, A. P. Arkin, FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS ONE* **5**, e9490 (2010).

43. A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).

44. G. P. Fournier, J. P. Gogarten, Rooting the ribosomal tree of life. *Mol. Biol. Evol.* **27**, 1792–1801 (2010).

45. N. Lartillot, T. Lepage, S. Blanquart, PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* **25**, 2286–2288 (2009).

46. J. L. Thorne, H. Kishino, I. S. Painter, Estimating the rate of evolution of the rate of molecular evolution. *Molecular Biology and Evolution* **15**, 1647–1657 (1998).

47. A. J. Drummond, S. Y. W. Ho, M. J. Phillips, A. Rambaut, Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4**, e88 (2006).

48. T. Lepage, D. Bryant, H. Philippe, N. Lartillot, A general comparison of relaxed molecular clock models. *Mol. Biol. Evol.* **24**, 2669–2680 (2007).

49. M. Kanehisa, M. Furumichi, Y. Sato, M. Ishiguro-Watanabe, M. Tanabe, KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res.* **49**, D545–D551 (2021).

50. M. Kanehisa, Toward understanding the origin and evolution of cellular organisms. *Protein Sci.* **28**, 1947–1951 (2019).

51. M. Kanehisa, S. Goto, KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).

52. K. Mendler, *et al.*, AnnoTree: visualization and exploration of a functionally annotated microbial tree of life. *Nucleic Acids Res.* **47**, 4442–4448 (2019).

53. R. C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

54. R. Sanchez, *et al.*, Phylemon 2.0: a suite of web-tools for molecular evolution, phylogenetics,

phylogenomics and hypotheses testing. *Nucleic Acids Research* **39**, W470–W474 (2011).

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

56. B. Q. Minh, *et al.*, IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).

57. E. Jacox, C. Chauve, G. J. Szöllősi, Y. Ponty, C. Scornavacca, ecceTERA: comprehensive gene tree-species tree reconciliation using parsimony. *Bioinformatics* **32**, 2056–2058 (2016).

58. W. Duchemin, *et al.*, RecPhyloXML: a format for reconciled gene trees. *Bioinformatics* **34**, 3646–3652 (2018).

59. L. A. David, E. J. Alm, Rapid evolutionary innovation during an Archaean genetic expansion. *Nature* **469**, 93–96 (2011).

60. K. Raymann, C. Brochier-Armanet, S. Gribaldo, The two-domain tree of life is linked to a new root for the Archaea. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 6670–6675 (2015).

61. T. A. Williams, C. J. Cox, P. G. Foster, G. J. Szöllősi, T. M. Embley, Phylogenomics provides robust support for a two-domains tree of life. *Nat Ecol Evol* **4**, 138–147 (2020).

62. K. Zaremba-Niedzwiedzka, *et al.*, Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* **541**, 353–358 (2017).

63. S. J. Mojzsis, *et al.*, Evidence for life on Earth before 3,800 million years ago. *Nature* **384**, 55–59 (1996).

64. J. L. Eigenbrode, K. H. Freeman, Late Archean rise of aerobic microbial ecosystems. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 15759–15764 (2006).

65. J. M. Wolfe, G. P. Fournier, Horizontal gene transfer constrains the timing of methanogen evolution. *Nat Ecol Evol* **2**, 897–903 (2018).

66. A. Bekker, *et al.*, Dating the rise of atmospheric oxygen. *Nature* **427**, 117–120 (2004).

67. P. Sánchez-Baracaldo, G. Bianchini, J. D. Wilson, A. H. Knoll, Cyanobacteria and biogeochemical cycles through Earth history. *Trends Microbiol.* (2021) https:/doi.org/10.1016/j.tim.2021.05.008.

68. K. Pang, *et al.*, The nature and origin of nucleus-like intracellular inclusions in Paleoproterozoic eukaryote microfossils. *Geobiology* **11**, 499–510 (2013).

69. E. J. Javaux, C. P. Marshall, A. Bekker, Organic-walled microfossils in 3.2-billion-year-old shallow-marine siliciclastic deposits. *Nature* **463**, 934–938 (2010).

70. T. M. Gibson, *et al.*, Precise age of Bangiomorpha pubescens dates the origin of eukaryotic photosynthesis. *Geology* **46**, 135–138 (2018).

71. N. J. Butterfield, Bangiomorpha pubescens n. gen., n. sp.: implications for the evolution of sex, multicellularity, and the Mesoproterozoic/Neoproterozoic radiation of eukaryotes. *Paleobiology* **26**, 386–404 (2000).

72. K. Pang, *et al.*, Nitrogen-Fixing Heterocystous Cyanobacteria in the Tonian Period. *Curr. Biol.* **28**, 616–622.e1 (2018).

73. S. Golubic, V. N. Sergeev, A. H. Knoll, Mesoproterozoic Archaeoellipsoidès: akinetes of heterocystous cyanobacteria. *Lethaia* **28**, 285–298 (1995).

74. G. D. Love, *et al.*, Fossil steroids record the appearance of Demospongiae during the Cryogenian period. *Nature* **457**, 718–721 (2009).

75. C. Burke, P. Steinberg, D. Rusch, S. Kjelleberg, T. Thomas, Bacterial community assembly based on functional genes rather than species. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 14288–14293 (2011).

76. S. Sievert, R. Kiene, H. Schulz-Vogt, The Sulfur Cycle. *Oceanography* **20**, 117–123 (2007).

77. C. Dahl, *et al.*, Genes involved in hydrogen and sulfur metabolism in phototrophic sulfur bacteria. *FEMS Microbiol. Lett.* **180**, 317–324 (1999).

78. C. G. Friedrich, F. Bardischewsky, D. Rother, A. Quentmeier, J. Fischer, Prokaryotic sulfur oxidation. *Current Opinion in Microbiology* **8**, 253–259 (2005).

79. M. Wagner, A. J. Roger, J. L. Flax, G. A. Brusseau, D. A. Stahl, Phylogeny of dissimilatory sulfite reductases supports an early origin of sulfate respiration. *J. Bacteriol.* **180**, 2975–2982 (1998).

80. B. Meyer, J. Kuever, Phylogeny of the alpha and beta subunits of the dissimilatory adenosine-5'-phosphosulfate (APS) reductase from sulfate-reducing prokaryotes – origin and evolution of the dissimilatory sulfate-reduction pathway. *Microbiology* **153**, 2026–2044 (2007).

81. B. Meyer, J. F. Imhoff, J. Kuever, Molecular analysis of the distribution and phylogeny of the soxB gene among sulfur-oxidizing bacteria - evolution of the Sox sulfur oxidation enzyme system. *Environ. Microbiol.* **9**, 2957–2977 (2007).

82. U. Zander, *et al.*, Structural basis for the oxidation of protein-bound sulfur by the sulfur cycle molybdohemo-enzyme sulfane dehydrogenase SoxCD. *J. Biol. Chem.* **286**, 8349–8360 (2011).

83. C. L. Blättler, *et al.*, Two-billion-year-old evaporites capture Earth's great oxidation. *Science* **360**, 320–323 (2018).

84. B. B. Jørgensen, A. J. Findlay, A. Pellerin, The Biogeochemical Sulfur Cycle of Marine Sediments. *Front. Microbiol.* **10**, 849 (2019).

85. M. R. Warke, *et al.*, The Great Oxidation Event preceded a Paleoproterozoic "snowball Earth." *Proceedings of the National Academy of Sciences* **117**, 13314–13320 (2020).

86. J. S. Boden, K. O. Konhauser, L. J. Robbins, P. Sánchez-Baracaldo, Timing the evolution of antioxidant enzymes in cyanobacteria. *Nat. Commun.* **12**, 4742 (2021).

87. W. Ghosh, B. Dam, Biochemistry and molecular biology of lithotrophic sulfur oxidation by taxonomically and ecologically diverse bacteria and archaea. *FEMS Microbiol. Rev.* **33**, 999–1043 (2009).

88. G. Feulner, C. Hallmann, H. Kienert, Snowball cooling after algal rise. *Nature Geoscience* **8**, 659–662 (2015).

89. R. J. Charlson, J. E. Lovelock, M. O. Andreae, S. G. Warren, Oceanic phytoplankton, atmospheric sulphur, cloud albedo and climate. *Nature* **326**, 655–661 (1987).

90. J. Farquhar, J. Savarino, S. Airieau, M. H. Thiemens, Observation of wavelength-sensitive mass-independent sulfur isotope effects during SO_2photolysis: Implications for the early atmosphere. *Journal of*

Geophysical Research: Planets **106**, 32829–32839 (2001).

91. I. Halevy, Production, preservation, and biological processing of mass-independent sulfur isotope fractionation in the Archean surface environment. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 17644–17649 (2013).

92. Y. Ueno, *et al.*, Geological sulfur isotopes indicate elevated OCS in the Archean atmosphere, solving faint young sun paradox. *Proceedings of the National Academy of Sciences* **106**, 14784–14789 (2009).

93. E. P. Reeves, J. M. McDermott, J. S. Seewald, The origin of methanethiol in midocean ridge hydrothermal fluids. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 5474–5479 (2014).

94. W. Martin, M. J. Russell, On the origin of biochemistry at an alkaline hydrothermal vent. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **362**, 1887–1925 (2007).

95. M. D. Schulte, K. L. Rogers, Thiols in hydrothermal solution: standard partial molal properties and their role in the organic geochemistry of hydrothermal environments. *Geochimica et Cosmochimica Acta* **68**, 1087–1097 (2004).

96. S. D. Domagal-Goldman, V. S. Meadows, M. W. Claire, J. F. Kasting, Using biogenic sulfur gases as remotely detectable biosignatures on anoxic planets. *Astrobiology* **11**, 419–441 (2011).

Tables

Table 1. Fossil calibration points used for calibrating molecular clocks. Calibration points were set as the hard constraint in Phylobayes, indicating the latest date by which a specific clade must have split. The "conservative" time points reflect the dates for which there is the most consensus; "liberal" time points reflect the earliest date reported in the literature.

Calibration Events	Conservative (Ga)	Liberal (Ga)
LUCA (set as root prior)	>3.8 (63)	>3.8 (63)
Origin of methanogenesis	>2.7 (64)	>3.51 (65)
Origin of oxygenic photosynthesis	>2.45 (66)	>3.0 (20) (67)
Origin of eukaryotes	>1.7 (68)	>3.2 (69)
Origin of plastids/Rhodophytes diverge	>1.05 (70)	>1.5 (71)
Akinetes diverge from cyanobacteria lacking cell differentiation	>1.0 (72)	>1.5 (73)
Origin of animals	>0.635 (74)	>0.635 (74)

Table 2. Sulfur cycling genes analyzed in this study. Number of gene hits indicates the number of genes identified among the genomes included in the species tree; IQ-Tree Model indicates the model of evolution used for generating the gene tree as determined by IQ-Tree; number of bootstraps indicates the number of bootstraps used for reconciling the gene tree with the species tree; number of loss, duplication, and HGT events are reported as determined by ecceTERA using the CIR clock model.

Gene	KEGG Orthology number	Number of gene hits	Metabolic pathway	IQ-Tree Model	Number of Bootstraps	Loss events	Dupli-cation events	HGT events	Total number of events
aprA	K00394	110	Dissimilatory sulfate oxidation/ reduction	LG+I+G4	800 (converged)	40	4	79	123
aprB	K00395	103	Dissimilatory sulfate oxidation/ reduction	LG+I+G4	800 (converged)	46	3	77	126
dsrA	K11180	83	Dissimilatory sulfate oxidation/ reduction	LG+I+G4	650 (converged)	37	2	50	89

dsrB	K11181	84	Dissimilatory sulfate oxidation/ reduction	LG+I+G4	1200 (converged)	29	2	28	79
soxA	K17222	40	Thiosulfate oxidation/ reduction	LG+I+G4	2000 (converged)	33	3	16	52
soxB	K17224	45	Thiosulfate oxidation/ reduction	LG+I+G5	650 (converged)	33	4	23	60
soxC	K17225	105	Thiosulfate oxidation/ reduction	LG+I+G6	3000 (converged)	32	7	78	117
soxX	K17223	38	Thiosulfate oxidation/ reduction	LG+I+G7	4000 (converged)	21	3	18	42
soxY	K17226	38	Thiosulfate oxidation/ reduction	LG+I+G8	6000 (converged)	17	2	19	38
soxZ	K17227	58	Thiosulfate oxidation/ reduction	LG+G4	1000 (not converged)	32	9	26	67
dmdA	K17486	28	Volatile organic sulfur cycling	LG+G4	2450 (converged)	1	1	24	26
dmsA	K07306	92	Volatile organic sulfur cycling	LG+G4	450 (converged)	5	11	75	91
mddA	K21310	65	Volatile organic sulfur cycling	LG+F+I+G4	2150 (converged)	11	3	52	66

Table 3. Identification of gene loss, duplication, and horizontal gene transfer events according to reconciliation with the chronogram generated using the CIR clock model and conservative calibration points, as identified by ecceTERA. Also shown is the gene birth date, defined here as the latest possible timing for the gene birth based on the earliest event identified for that gene according to ecceTERA.

Gene	**Loss**	**Duplica-tion**	**HGT**	**Total number of events**	**Birth Date Range (mya)**	**Midpoint Birth Date (mya)**
aprA	40	4	79	123	2991.975 - 3304.41	3148.192
aprB	46	3	77	126	2894.515 - 3088.86	2991.687

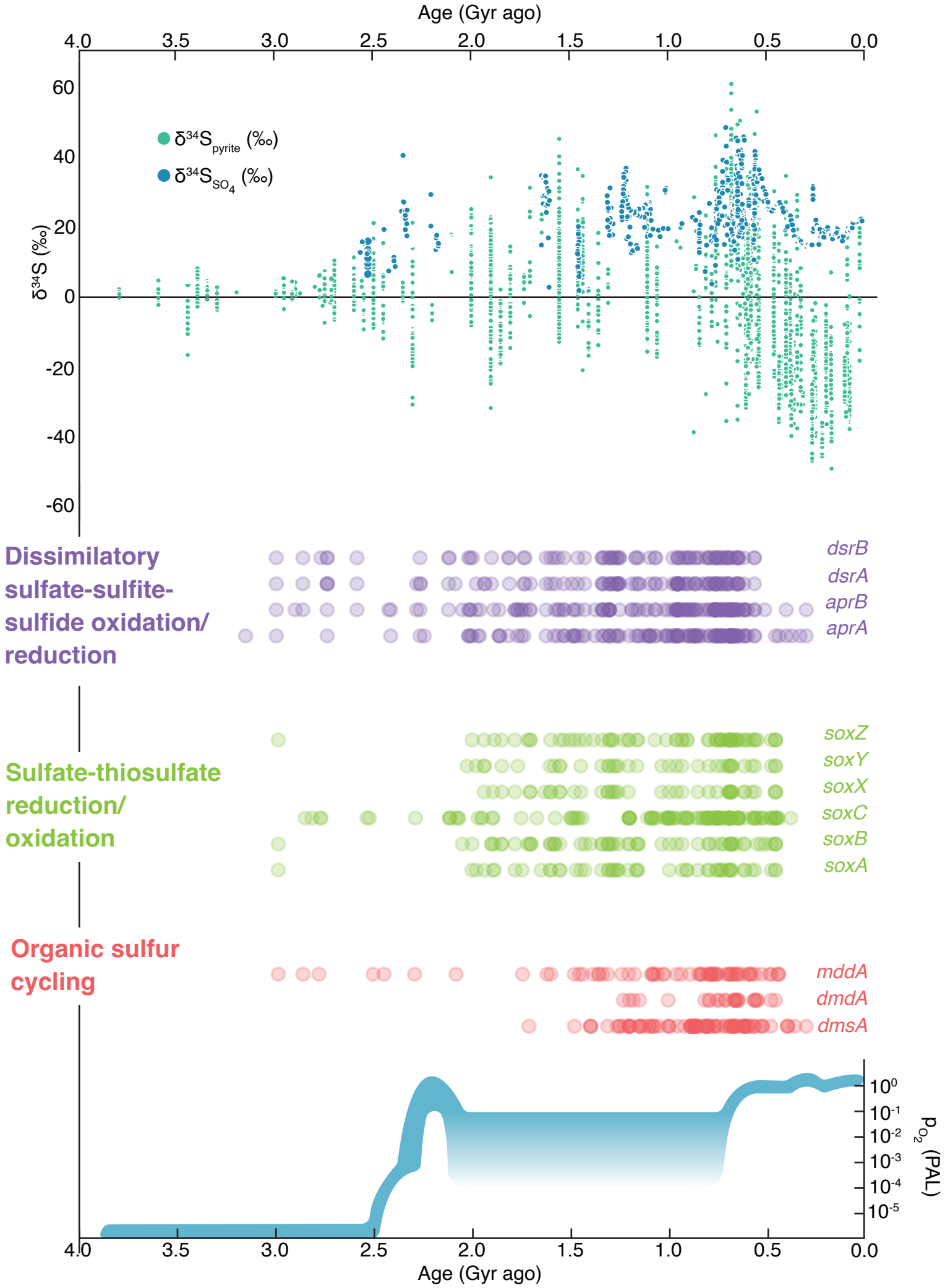
dsrA	37	2	50	89	2894.515 - 3088.86	2991.687
dsrB	29	2	28	79	2894.515 - 3088.86	2991.687
soxA	33	3	16	52	2902.415 - 3060.56	2981.487
soxB	33	4	23	60	2902.415 - 3060.56	2981.487
soxC	32	7	78	117	2287.77 - 3400.795	2844.282
soxX	21	3	18	42	1882.185 - 1981.1	1931.642
soxY	17	2	19	38	1968.43 - 2070.21	2019.32
soxZ	32	9	26	67	2902.415 - 3060.56	2981.487
dmdA	1	1	24	26	1106.7795 - 1337.14	1221.9597
dmsA	5	11	75	91	1650.365 - 1757.72	1704.042
mddA	11	3	52	66	2902.415 - 3060.56	2981.487

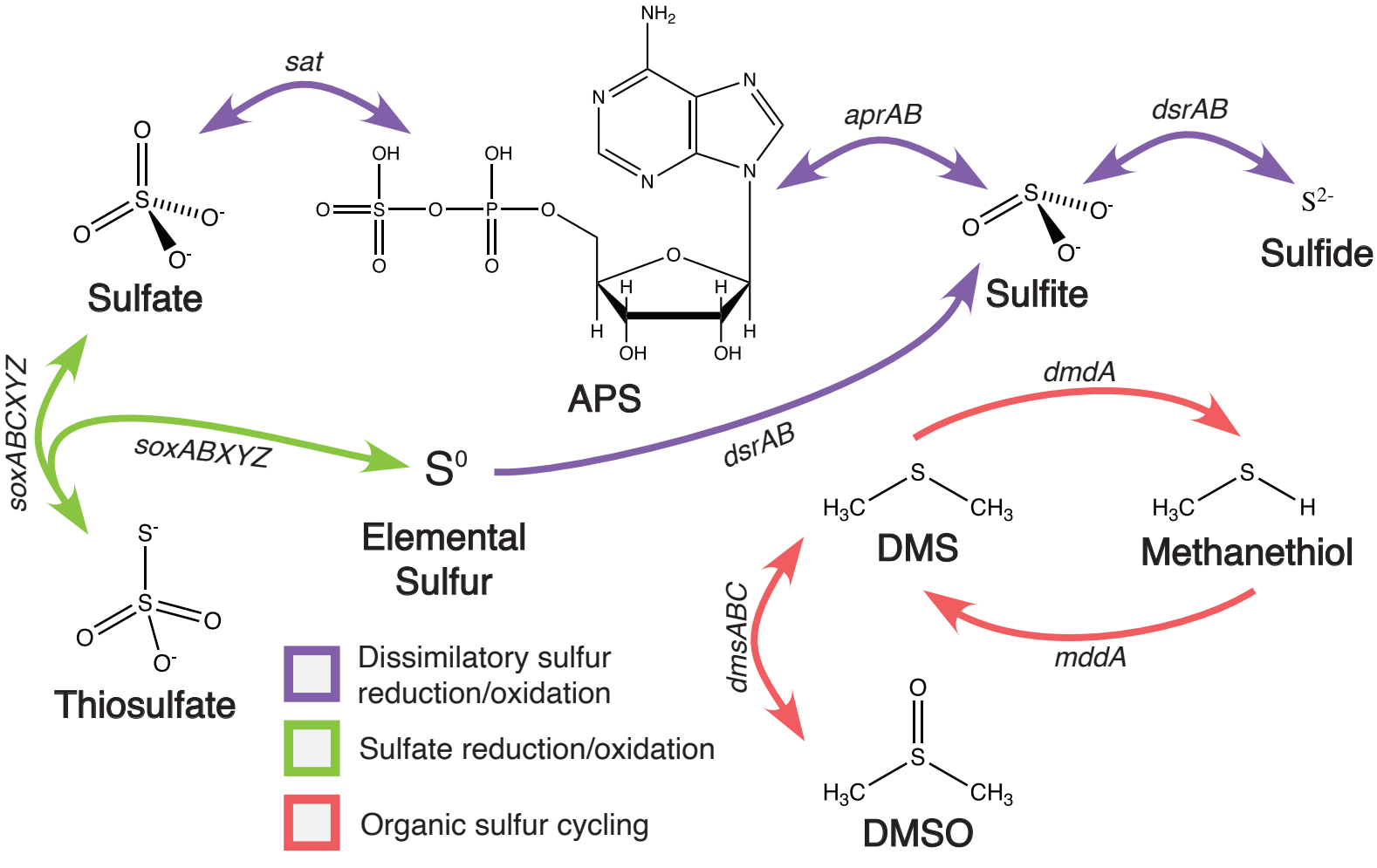
Figure Legends:

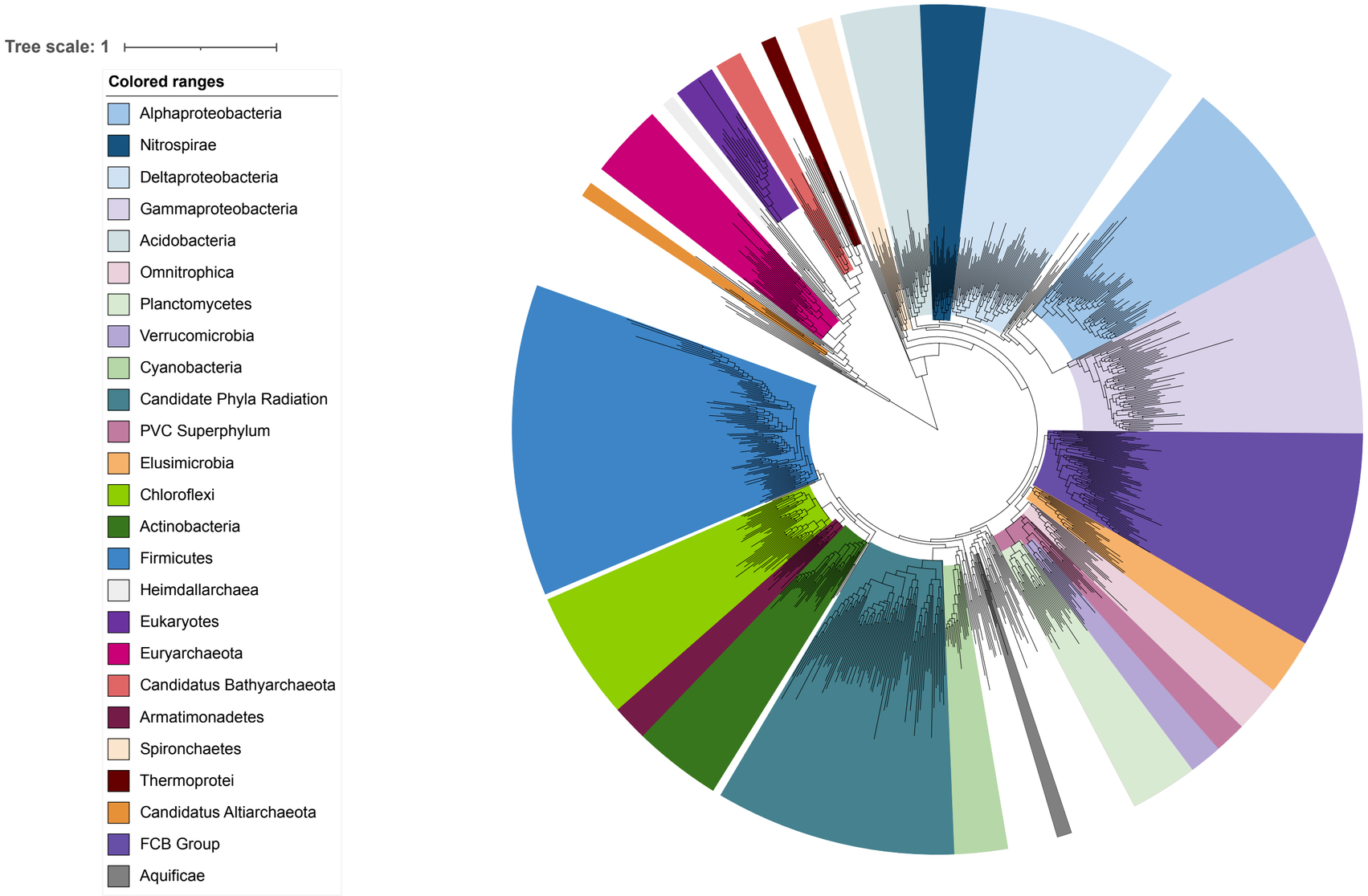
Figure 1. Schematic of the biological sulfur cycle, highlighting the genes included in this analysis (note that *sat* was excluded from the analysis because it is also used in several other pathways).

Figure 2. Tree of life ("species tree") used for this study. The tree includes 871 genomes in total, including 777 bacterial, 80 archaeal, and 14 eukaryotic genomes. The bacterial and archaeal genomes represent one genome per order based on the GTDB taxonomy.

Figure 3. Comparison of gene reconciliation results with previously published geochemistry data. The data represented here were calculated using the CIR clock model and conservative calibration points. The graph above depicting isotopic data was adapted from Fike et al. 2015; the graph below depicting oxygen concentrations was adapted from Lyons et al. 2014.


Age (Gyr ago)
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
60
40
20
0
-20
-40
-60
δ34S (‰)
δ34Spyrite (‰)
δ34SSO4 (‰)
Dissimilatory sulfate-sulfite-sulfide oxidation/ reduction
dsrB
dsrA
aprB
aprA
Sulfate-thiosulfate reduction/ oxidation
soxZ
soxY
soxX
soxC
soxB
soxA
Organic sulfur cycling
mddA
dmdA
dmsA
10^0
10^-1
10^-2
10^-3
10^-4
10^-5
pO2 (PAL)
Age (Gyr ago)


sat
Sulfate
APS
aprAB
dsrAB
Sulfite
Sulfide
soxABCXYZ
soxABXYZ
S0
Elemental Sulfur
dsrAB
Thiosulfate
dmdA
DMS
Methanethiol
dmsABC
mddA
DMSO
Dissimilatory sulfur reduction/oxidation
Sulfate reduction/oxidation
Organic sulfur cycling


Tree scale: 1
Colored ranges
Alphaproteobacteria
Nitrospirae
Deltaproteobacteria
Gammaproteobacteria
Acidobacteria
Omnitrophica
Planctomycetes
Verrucomicrobia
Cyanobacteria
Candidate Phyla Radiation
PVC Superphylum
Elusimicrobia
Chloroflexi
Actinobacteria
Firmicutes
Heimdallarchaea
Eukaryotes
Euryarchaeota
Candidatus Bathyarchaeota
Armatimonadetes
Spironchaetes
Thermoprotei
Candidatus Altiarchaeota
FCB Group
Aquificae