Title: Dramatic expansion of microbial groups that shape the global sulfur cycle

Authors: Karthik Anantharaman[1*], Sean P. Jungbluth[2], Rose S. Kantor[3], Adi Lavy[1], Lesley A. Warren[4], Michael S. Rappé[5], Brian C. Thomas[1], and Jillian F. Banfield[1,6,7*]

Affiliations:

[1]Department of Earth and Planetary Sciences, Berkeley, CA, USA

[2]DOE Joint Genome Institute, Walnut Creek, CA, USA

[3]Department of Plant and Microbial Biology, University of California, Berkeley, CA, USA

[4]Department of Civil Engineering, University of Toronto, Ontario, CANADA

[5]Hawaii Institute of Marine Biology, University of Hawaii at Manoa, Kaneohe, HI, USA

[6]Department of Environmental Science, Policy, and Management, Berkeley, CA, USA

[7]Earth and Environmental Sciences, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

*Corresponding author

Email: Karthik.a@berkeley.edu

Address: 307 McCone Hall, Berkeley, CA 94720

Abstract

The biogeochemical cycle of sulfur on Earth is driven by microbial sulfate reduction, yet organisms from relatively few lineages have been implicated in this process. Recent studies using functional marker genes have detected abundant, novel dissimilatory sulfite reductases that confer the capacity for microbial sulfate reduction and could do not be affiliated with known organisms. Thus, the identity of a significant fraction of sulfate reducing microbes has remained elusive. Here we report the discovery of the capacity for sulfate reduction in the genomes of organisms from twelve bacterial and archaeal phyla, thereby doubling the number of microbial phyla associated with this process. Eight of the twelve newly identified groups are candidate phyla that lack isolated representatives, a finding only possible given genomes from metagenomes. Two candidate phyla, *Candidatus* Rokubacteria and *Candidatus* Hydrothermarchaeota contain the earliest evolved genes. The capacity for sulfate reduction has been laterally transferred in multiple events within some phyla, and a key gene potentially capable of switching sulfur oxidation to sulfate reduction in associated cells has been acquired by putatively symbiotic bacteria. We conclude that functional predictions based on phylogeny will significantly underestimate the extent of sulfate reduction across Earth's ecosystems. Understanding the prevalence of this capacity is integral to interpreting the carbon cycle because sulfate reduction is often coupled to turnover of buried organic carbon. Our findings expand the diversity of microbial groups associated with sulfur transformations in the environment and motivate revision of biogeochemical process models based on microbial community composition.

The cycling of sulfur is one of Earth's major biogeochemical processes. Sulfate reduction may be an early evolved microbial metabolism, given evidence for biological fractionation of sulfur isotopes around 3.5 billion years ago[1], and it remains an important metabolic platform for anaerobic life[2]. In natural ecosystems, human microbiomes and engineered systems, this process is important because the product hydrogen sulfide (H_2S) is toxic[3], can corrode steel[4], and sour oil reservoirs[5]. Overall, sulfate reduction is a primary driver in the carbon cycle, and is responsible for conversion of ~30% of the organic carbon flux to CO_2 in sedimentary environments[6]. Importantly, the coupling of sulfate reduction to oxidation of H_2, small chain fatty acids or other carbon compounds limits the availability of these substrates to other organisms and alters the energetics via syntrophic interactions[7]. All of these processes also impact methane production. Given the many reasons why the biological conversion of sulfate/sulfite to sulfide is important, it is vital that we understand which organisms can carry out the reactions and the pathways involved.

Dissimilatory sulfite reductase (dsr) genes confer bacteria and archaea the ability to grow via reduction of sulfite and can function in reverse in some organisms that oxidize sulfur[8,9]. The phylogenetic distribution of organisms with dsr genes has been considered to be quite limited[10]. The recent availability of thousands of genomes from organisms belonging to many newly

sampled phyla has provided the opportunity to test for the presence of dsr genes in bacteria and archaea that have not previously been associated with dissimilatory sulfur metabolism[11].

Results and Discussion

We analyzed genomes reconstructed from metagenomic sequence datasets recovered from five distinct terrestrial and marine subsurface environments. The sampling sites included an aquifer adjacent to the Colorado River, USA[12], a deep subsurface CO_2 geyser in Utah, USA[13], a deep borehole in Japan[14], an acidic sulfide mine waste rock site in Canada[15], and deep subseafloor basaltic crustal fluids of the hydrothermally active Juan de Fuca ridge flank in the Pacific Ocean[16]. We identified dsr genes in 122 near-complete microbial genomes (**Supplementary Table 1**). Phylogenetic analyses using a set of 16 concatenated ribosomal proteins (RP) and the small subunit ribosomal (SSU) RNA gene show that these genomes belong to organisms from 16 distinct phylum-level lineages (**Table 1**), 12 of which were not known to have dsr genes[10]. In addition, we identified anaerobic sulfite reductase (asr) genes required for sulfite reduction in two bacterial groups not previously reported to have this capacity[17]. All of the identified catalytic proteins (DsrA, DsrB, and AsrC) contained all conserved sulfite reductase residues and secondary structure elements for the formation of α helices and β sheets[18] (**Supplementary Fig. 1, Supplementary Fig. 2, Supplementary Fig. 3**).

Given our interest in identifying organisms with the capacity to produce sulfide, we searched genomes for operons that contained genes encoding DsrD. This gene is considered a marker for sulfate reduction because it is absent in bacteria that use the reverse dissimilatory sulfate reduction (reverse-dsr) pathway for sulfur oxidation[19]. Although the exact function of the DsrD protein is unclear, the presence of winged-helix domains in its structure and its association with other core proteins of the dsr complex (*dsrABC*) suggest a regulatory role in bacterial sulfate reduction[20]. We identified 78 genomes that encode at least *dsrABCD* (**Supplementary Fig. 4**). A multiple alignment of DsrD sequences confirmed highly conserved residues, indicating that the proteins are likely active (**Supplementary Fig. 5**). These putative sulfate/sulfite reducing microorganisms affiliate with eight distinct phyla not previously reported to be capable of these processes. Four are phyla with isolated representatives (*Acidobacteria*, *Armatimonadetes*, *Ignavibacteria*, *Planctomycetes*) and four are considered candidate phyla due to the absence of isolated representatives (*Candidatus* Zixibacteria, *Candidatus* Schekmanbacteria, *Candidatus* Desantisbacteria, *Candidatus* Lambdaproteobacteria) (**Fig. 1**). The asr pathway for sulfite reduction was found in members of two candidate phyla, *Candidatus* Omnitrophica and *Candidatus* Riflebacteria (**Supplementary Fig. 6**).

Surprisingly, we identified *dsrD* genes in eight genomes of organisms affiliating with *Candidatus* Falkowbacteria, putatively symbiotic bacteria within the Parcubacteria superphylum of the candidate phyla radiation (CPR)[21]. There is no indication of the presence of other dsr

genes in these genomes. Given the predicted close physical and metabolic interactions between CPR and their hosts, we suggest that this small protein could augment host metabolism, as sometimes occurs with viruses/phage and their hosts[22]. CPR are common in aquifers where conditions oscillate between oxic and anoxic[12]. Potentially, populations with the *dsrD* gene could maintain host function by enabling switching between sulfur oxidation to sulfate reduction as conditions change. The predicted Falkowbacteria DsrD protein sequences cluster with sequences from well characterized *Deltaproteobacteria* capable of sulfate reduction, suggesting that these CPR may have acquired the gene by lateral gene transfer (LGT) from this group (**Supplementary Fig. 7**).

Prior analyses have suggested the LGT events involving the catalytic dsr subunits A and B genes have occurred, but infrequently[9,10,23]. We used a concatenated *dsrAB* protein tree to reevaluate the extent to which LGT has influenced the organismal distribution of these genes (**Fig. 1**). We found that organism phylogeny is not a reliable predictor of the grouping of these sequences. In fact, phylogenetic evidence suggests that LGT events have introduced these genes into some phyla in multiple independent events (e.g., Nitrospirae sequences place in five distinct locations on the tree). However, almost all organisms lacking *dsrD* genes cluster together with organisms known to be sulfur oxidizers in the *dsrAB* tree. Based on this clustering, the group implicated in elemental sulfur oxidation now includes bacteria from three phyla: *Nitrospirae*, *Nitrospinae*, and *Candidatus* Muproteobacteria (**Fig. 1, Fig. 2**). Importantly, organisms from two candidate phyla, *Candidatus* Rokubacteria and *Candidatus* Hydrothermarchaeota lack *dsrD* genes but their *dsrAB* sequences cluster with 'reductive archaeal-type' dsr sequences found in thermophilic sulfite/thiosulfate-reducing *Crenarchaeota* and *Aigarchaeota*. The branch that includes sequences from *Candidatus* Rokubacteria, *Candidatus* Hydrothermarchaeota, *Crenarchaeota* and *Aigarchaeota* is basal within the *dsrAB* tree (**Fig. 1**). The phylogenetic placements of *Candidatus* Rokubacteria and *Candidatus* Hydrothermarchaeota with sulfite/thiosulfate-reducing archaea implicates these organisms in either sulfate or sulfite reduction.

To determine whether the organisms reduce sulfate vs. sulfite to sulfide we looked for the genes involved in the reduction of sulfate to sulfite, specifically adenosine phosphosulfate reductase subunits A, B (*aprAB*), sulfate adenylyl transferase (*sat*), and quinone-interacting membrane-bound oxidoreductase subunits A, B, C (*qmoABC*)[24–26]. Unlike *Crenarchaeota* and *Aigarchaeota* that can only reduce sulfite, *Candidatus* Rokubacteria have apr, sat, and the qmo genes that are required to gain energy from reduction of sulfate to sulfite. Surprisingly, phylogenetic analyses show that the *Candidatus* Rokubacteria and *Candidatus* Hydrothermarchaeota qmo (and *apr* and *sat*) genes cluster with sequences from other sulfate-reducing bacteria (**Supplementary Fig. 8, Supplementary Fig. 9, Supplementary Fig. 10**). Thus, we suggest that *Candidatus* Rokubacteria have a system for sulfate reduction to sulfide of hybrid origin, with ancient DsrA and B genes related to those found in archaea, and other components similar to bacterial sequences (**Fig. 3**).

Given the lack of *dsrD* in *Candidatus* Rokubacteria and *Candidatus* Hydrothermarchaeota, we sought evidence for hypothetical genes in proximity to dsr genes that may be markers for the sulfate/sulfite reduction pathway. We identified a hypothetical gene that encodes for the N-terminal domain of an anti-sigma factor antagonist protein[27] that almost always occurs within the operon encoding dsr genes (**Fig. 2**). This hypothetical protein is part of a protein family that includes the *Bacillus subtitlis* RsbT co-antagonist protein rsbRD, which are important components of the stressosome and function as negative regulators of the general stress transcription factor sigma-B[28]. This gene is unique to sulfide-producing organisms and is absent in sulfur-oxidizing organisms (except for the *Chlorobiae* clade) (**Supplementary Fig. 11**). The gene always precedes the electron transport components encoded by *dsrMKJOP* genes and is fused with *dsrM* in some organisms (**Supplementary Fig. 12**). From structural predictions and conserved motifs, we hypothesize that it likely performs a regulatory function (**Supplementary Fig. 13**). We refer to this newly identified gene as 'dsrX' and suggest that it may serve as an additional marker for genome-based prediction of sulfate or sulfite reducing metabolism.

In order to understand the energy metabolism and ecology of these novel sulfate reducing organisms, we investigated potential electron donors for sulfate/sulfite reduction. Specifically, we targeted genes involved in the oxidation of hydrogen[29] (Ni-Fe hydrogenase groups I, IIa, IIb, IIIa, IIIb, IIIc, IIId) and transformation of organic carbon compounds (Genes involved in breakdown of Cellulose, Hemicellulose, Chitin, Pectin, Starch, Amino sugars, Other monosaccharides and polysaccharides)[30]. Our analyses show that organisms from 10 of the 12 sulfate reducing lineages possess the ability to utilize hydrogen as an electron donor for sulfate reduction (**Supplementary Table 2).** On the other hand, organisms from all 12 lineages possessed the ability to breakdown organic compounds although the diversity of genes encoding for specific carbohydrate active enzymes varied greatly across phyla (**Supplementary Table 3**). We propose that sulfate reduction by organisms from these newly identified lineages likely serves an important control on the transformation of organic carbon in the terrestrial and marine subsurface.

By the Proterozoic Eon, sulfate reduction had become a significant biological process in the oceans[31,32]. Based on phylogenomic arguments and isotopic records, it was suggested that the capacity to reduce sulfite to sulfide emerged in thermophilic archaea around 3.5 billion years ago, and that mesophilic sulfate reducers evolved only after the rise in atmospheric oxygen level[1,33]. Our findings indicate a complex evolutionary history of this capacity involving extensive LGT of dsr genes. Consequently, it may be impossible to constrain the lineage in which this metabolism first appeared. The ability to reduce sulfate/sulfite is now predicted in a much wider diversity of mesophilic bacterial and archaeal groups than was recognized previously. We conclude that many groups of microorganisms now known to have genes involved in dissimilatory sulfur metabolism impact biogeochemical processes in marine and

196 terrestrial sediments, aquifers, wetlands, methane seeps, coastal marshes and estuaries, as well as
197 agricultural and human microbiomes. Many are organisms from well-studied phyla, but still
198 novel at the Genus to Class levels, but others are organisms from candidate phyla known only
199 based on their genomes. The results underline the value of genomic analyses for prediction of
200 key ecosystem capacities that cannot be made based on rRNA gene surveys and motivate
201 targeted cultivation strategies for organisms currently lacking laboratory tractable
202 representatives.
203

204 ## Methods

205

206 **Sample Collection and Data Processing.** Details of sample collection (Sampling, DNA
207 Extraction), individual sample geochemical measurements, and data processing (DNA
208 sequencing, Assembly, Annotation, Binning, Genome Completion Estimates) are described in
209 detail elsewhere[12–16].
210

211 **Identification of sulfate reducing organisms.** Genome-specific metabolic potential for sulfate
212 reduction was determined in an iterative manner by (A) Searching all predicted ORFs in a
213 genome with hmm profiles for *dsrA* and *dsrB* from TIGRfam[34], and *dsrD* from Pfam[35] using
214 hmmscan v3.1b2[36], and (B) Generation of custom hmm profiles for *dsrA*, *dsrB* and *dsrD* using
215 hits generated from step (A) and searching all predicted ORFs again for the above genes. For
216 generation of custom HMM profiles, reference sequences and identified genes from step (A)
217 were aligned using MUSCLE v3.8.31[37] with default parameters followed by manually trimming
218 the start and ends of the alignment. The alignment was converted into Stockholm format and
219 databases were built using hmmscan[36]. Individual noise and trusted cutoffs for all HMMs were
220 determined by manual inspection and are built into the custom HMM profiles.
221

222 **Sequence alignment and phylogeny.** Phylogenetic analyses were performed as follows:
223

224 Each individual gene (*dsrA*, *dsrB*, *dsrC*, *dsrD*, *aprA*, *aprB*, *dsrX*, *qmoA*, *qmoB*, *sat*) was aligned
225 along with reference sequences using MUSCLE[37] with default parameters. All alignments were
226 manually refined by trimming the start and ends and removing all columns with >95% gaps. For
227 generation of concatenated alignments (*dsrAB*, *qmoAB*, and *aprAB*), individual alignments were
228 concatenated in Geneious version 7[38]. In construction of the concatenated qmo tree, only
229 subunits A and B were used since subunit C is not universally present in sulfate reducing
230 organisms, being absent in sulfate reducing archaea. All phylogenetic analyses were inferred by
231 RAxML v8.0.26[39] implemented by the CIPRES Science Gateway[40]. RAxML was called as
232 follows:
233

234 For *asrABC*, *dsrD*, *sat*, *dsrX*, *qmoAB* trees:
235

raxmlHPC-PTHREADS -s input -N 1000 -n result -f a -p 12345 -x 12345 -m PROTGAMMAGTR.

For *dsrAB*, *aprAB* trees:

raxmlHPC-HYBRID -s input -N autoMRE -n result -f a -p 12345 -x 12345 -m PROTGAMMAGTR.

Conserved residues and motifs. Conserved residues and motifs in DsrA, DsrB, DsrC, AsrC, and DsrD proteins were identified by aligning the identified genes from all 122 genomes in this study with reference proteins[18,20,41]. Specific residues highlighted in **Supplementary Fig. 1**, **Supplementary Fig. 2**, **Supplementary Fig. 3**, **Supplementary Fig. 5**, and **Supplementary Fig. 14** were identified in *Desulfovibrio vulgaris*[42].

Structural models. We selected the DsrX proteins identified in *Desulfovibrio vulgaris* (WP_012611240) and Candidatus Rokubacteria CSP1-6 (KRT71371) for structural modeling. Protein models were predicted using the I-TASSER suite[43]. The models shown in **Supplementary Fig. 12** are the top predicted models out of the top five I-TASSER simulations. Both DsrX proteins used the identical top threading template from the sporulation inhibitor protein pXO1-118 from *Bacillus anthracis*[44].

Analyses of electron donors for sulfate reduction. Analyses of putative electron donors were centered around hydrogen and organic carbon compounds (carbohydrates). For identification of the potential for hydrogen oxidation, hmm searches were conducted by searching all predicted ORFs against individual HMM profiles for nickel-iron hydrogenases from Groups I, IIa, IIb, IIIa, IIIb, IIIc, and IIId. All hits above the noise cutoffs were inspected manually.

For identification of carbohydrate substrates for sulfate reduction, all predicted ORFs were searched against the CAZY HMM database[30]. Pre-filtering of hits was conducted using the following cutoffs: coverage: 0.40; e-value: 1e-18. To determine the specificity of enzymes, we established a set of 84 distinct reactions involving 189 enzyme families that allowed us to track specific substrates and products. All hits to Glycosyltransferases (GT) and Carbohydrate Binding Modules (CBM) were excluded from this analysis due to high incidence of false positives and/or difficulty in determining substrate specificity.

Data availability. NCBI Genbank, BioProject, BioSample, and Taxonomy ID (TaxID) accession numbers for individual genomes are listed in Supplemenatry Table 1. Genomes are also available through ggKbase: http://ggkbase.berkeley.edu/novel_sulfate_reducers (ggKbase is a 'live' site, genomes may be updated after publication). The JdFR-17, JdFR-18, and JdFR-19 genomes are also available through the Integrated Microbial Genomes and Microbiomes

database (IMG) through Genome IDs: 2728369317, 2728369320, 2728369322. Hmm databases used in this study are available from https://github.com/banfieldlab/metabolic-hmms. Additionally, ARB databases for *dsrA*, *dsrB*, and concatenated *dsrAB* genes are available from https://github.com/banfieldlab/dsrAB-ARB-db. The authors declare that all other data supporting the findings of this study are available within the article and its supplementary information files, or from the corresponding author on request.

References

1. Shen, Y., Buick, R. & Canfield, D. E. Isotopic evidence for microbial sulphate reduction in the early Archaean era. *Nature* **410,** 77–81 (2001).
2. Grein, F., Ramos, A. R., Venceslau, S. S. & Pereira, I. A. C. Unifying concepts in anaerobic respiration: Insights from dissimilatory sulfur metabolism. *Biochim. Biophys. Acta - Bioenerg.* **1827,** 145–160 (2013).
3. Huycke, M. M. & Gaskins, H. R. Commensal Bacteria, Redox Stress, and Colorectal Cancer: Mechanisms and Models. *http://dx.doi.org/10.1177/153537020422900702* (2016). doi:10.1177/153537020422900702
4. Stipaničev, M. *et al.* Corrosion behavior of carbon steel in presence of sulfate-reducing bacteria in seawater environment. *Electrochim. Acta* **113,** 390–406 (2013).
5. Gieg, L. M., Jack, T. R. & Foght, J. M. Biological souring and mitigation in oil reservoirs. *Appl. Microbiol. Biotechnol.* **92,** 263–282 (2011).
6. Bowles, M. W., Mogollón, J. M., Kasten, S., Zabel, M. & Hinrichs, K.-U. Global rates of marine sulfate reduction and implications for sub–sea-floor metabolic activities. *Science* **344,** (2014).
7. Pereira, I. A. C. *et al.* A comparative genomic analysis of energy metabolism in sulfate reducing bacteria and archaea. *Front. Microbiol.* **2,** 69 (2011).
8. Simon, J. & Kroneck, P. M. H. in *Adv. Microb. Physiol.* **62,** 45–117 (2013).
9. Loy, A. *et al.* Reverse dissimilatory sulfite reductase as phylogenetic marker for a subgroup of sulfur-oxidizing prokaryotes. *Environ. Microbiol.* **11,** 289–299 (2009).
10. Müller, A. L., Kjeldsen, K. U., Rattei, T., Pester, M. & Loy, A. Phylogenetic and environmental diversity of DsrAB-type dissimilatory (bi)sulfite reductases. *ISME J.* **9,** 1152–65 (2015).
11. Hug, L. A. *et al.* A new view of the tree of life. *Nat. Microbiol.* **1,** 16048 (2016).
12. Anantharaman, K. *et al.* Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat. Commun.* **13219,** (2016).
13. Probst, A. J. *et al.* Genomic resolution of a cold subsurface aquifer community provides metabolic insights for novel microbes adapted to high CO 2 concentrations. *Environ. Microbiol.* **19,** 459–474 (2017).
14. Hernsdorf, A. W. *et al.* Potential for microbial H2 and metal transformations associated with novel bacteria and archaea in deep terrestrial subsurface sediments. *ISME J.* (2017). doi:10.1038/ismej.2017.39
15. Kantor, R. S. *et al.* Lifting the veil on microbial sulfur biogeochemistry in mining wastewaters. **In Review,** (2017).
16. Jungbluth, S. P., Amend, J. P. & Rappé, M. S. Metagenome sequencing and 98 microbial genomes from Juan de Fuca Ridge flank subsurface fluids. *Sci. Data* **4,** 170037 (2017).

17. Hallenbeck, P. C., Clark, M. A. & Barrett, E. L. Characterization of anaerobic sulfite reduction by Salmonella typhimurium and purification of the anaerobically induced sulfite reductase. *J. Bacteriol.* **171,** 3008–15 (1989).
18. Dhillon, A., Goswami, S., Riley, M., Teske, A. & Sogin, M. Domain Evolution and Functional Diversification of Sulfite Reductases. *Astrobiology* **5,** 18–29 (2005).
19. Rabus, R. *et al.* in *Adv. Microb. Physiol.* **66,** 55–321 (2015).
20. Mizuno, N., Voordouw, G., Miki, K., Sarai, A. & Higuchi, Y. Crystal Structure of Dissimilatory Sulfite Reductase D (DsrD) Protein—Possible Interaction with B- and Z-DNA by Its Winged-Helix Motif. *Structure* **11,** 1133–1140 (2003).
21. Brown, C. T. *et al.* Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature* **523,** 208–211 (2015).
22. Anantharaman, K. *et al.* Sulfur Oxidation Genes in Diverse Deep-Sea Viruses. *Science* **344,** 757–760 (2014).
23. Klein, M. *et al.* Multiple Lateral Transfers of Dissimilatory Sulfite Reductase Genes between Major Lineages of Sulfate-Reducing Prokaryotes. *J. Bacteriol.* **183,** 6028–6035 (2001).
24. Meyer, B. & Kuever, J. 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).
25. Meyer, B., Kuever, J., Thamdrup, B., Canfield, D. & Simon, M. Homology Modeling of Dissimilatory APS Reductases (AprBA) of Sulfur-Oxidizing and Sulfate-Reducing Prokaryotes. *PLoS One* **3,** e1514 (2008).
26. Duarte, A. G., Santos, A. A. & Pereira, I. A. C. Electron transfer between the QmoABC membrane complex and adenosine 5′-phosphosulfate reductase. *Biochim. Biophys. Acta - Bioenerg.* **1857,** 380–386 (2016).
27. Kim, T.-J., Gaidenko, T. A. & Price, C. W. A Multicomponent Protein Complex Mediates Environmental Stress Signaling in Bacillus subtilis. *J. Mol. Biol.* **341,** 135–150 (2004).
28. Akbar, S. *et al.* New family of regulators in the environmental signaling pathway which activates the general stress transcription factor sigma(B) of Bacillus subtilis. *J. Bacteriol.* **183,** 1329–38 (2001).
29. Vignais, P. M. & Billoud, B. Occurrence, classification, and biological function of hydrogenases: an overview. *Chem. Rev.* **107,** 4206–72 (2007).
30. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42,** D490–D495 (2014).
31. Anbar, A. D. & Knoll, A. H. Proterozoic Ocean Chemistry and Evolution: A Bioinorganic Bridge? *Science* **297,** (2002).
32. Canfield, D. E., Habicht, K. S. & Thamdrup, B. The Archean sulfur cycle and the early history of atmospheric oxygen. *Science* **288,** 658–61 (2000).
33. Blank, C. E. Evolutionary timing of the origins of mesophilic sulphate reduction and oxygenic photosynthesis: a phylogenomic dating approach. *Geobiology* **2,** 1–20 (2004).
34. Haft, D. H., Selengut, J. D. & White, O. The TIGRFAMs database of protein families. *Nucleic Acids Res.* **31,** 371–3 (2003).
35. Finn, R. D. *et al.* Pfam: the protein families database. *Nucleic Acids Res.* **42,** D222–30 (2014).

36. Eddy, S. R. Accelerated Profile HMM Searches. *PLoS Comput. Biol.* **7,** e1002195 (2011).
37. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32,** 1792–7 (2004).
38. Kearse, M. *et al.* Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28,** 1647–9 (2012).
39. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30,** 1312–3 (2014).
40. Miller, M. A., Pfeiffer, W. & Schwartz, T. The CIPRES science gateway. in *Proc. 2011 TeraGrid Conf. Extrem. Digit. Discov. - TG '11* 1 (ACM Press, 2011). doi:10.1145/2016741.2016785
41. Venceslau, S. S., Stockdreher, Y., Dahl, C. & Pereira, I. A. C. The 'bacterial heterodisulfide' DsrC is a key protein in dissimilatory sulfur metabolism. *Biochim. Biophys. Acta - Bioenerg.* **1837,** 1148–1164 (2014).
42. Heidelberg, J. F. *et al.* The genome sequence of the anaerobic, sulfate-reducing bacterium Desulfovibrio vulgaris Hildenborough. *Nat. Biotechnol.* **22,** 554–559 (2004).
43. Yang, J. *et al.* The I-TASSER Suite: protein structure and function prediction. *Nat. Methods* **12,** 7–8 (2014).
44. Stranzl, G. R. *et al.* Structural Insights into Inhibition of Bacillus anthracis Sporulation by a Novel Class of Non-heme Globin Sensor Domains. *J. Biol. Chem.* **286,** 8448–8458 (2011).

Acknowledgements:

We thank David Burstein for inputs into hmm analysis and Christopher T. Brown for helpful discussion. This work was supported by Lawrence Berkeley National Laboratory's Sustainable Systems Scientific Focus Area funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under contract DE-AC02-05CH11231.

Additional information

Supplementary information is available for this paper

Correspondence and request for materials should be addressed to K.A., karthik.a@berkeley.edu

Competing interests

The authors declare no competing financial interests.

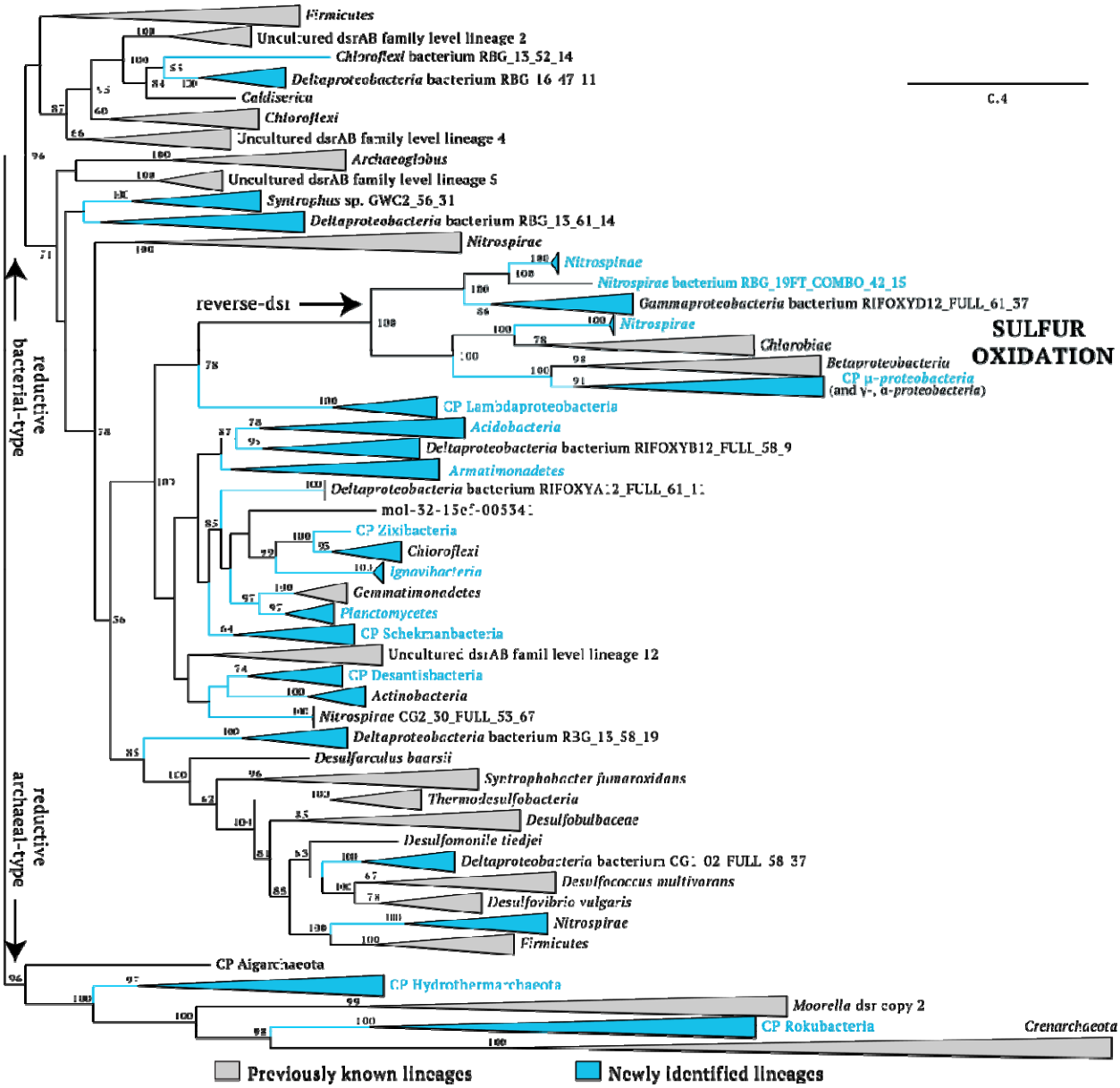

Fig.1 Phylogenetic tree of concatenated dsrAB proteins showing the diversity of organisms involved in dissimilatory sulfur cycling. Lineages in blue contain genomes reported in this study. Phylum-level lineages with first report of evidence for sulfur cycling are indicated by blue letters. Only bootstrap values >50 are shown. The complete tree is available with full bootstrap support values as Additional Data File S1.

Fig.2 Dsr operon structure in previously reported (black names) and newly reported groups (blue names). Interestingly, and in contrast to the previously studied organisms for which the operon is interrupted (=SS=), the entire dsr pathway (including electron transport chain and ancillary proteins) is often encoded in a single genomic region.

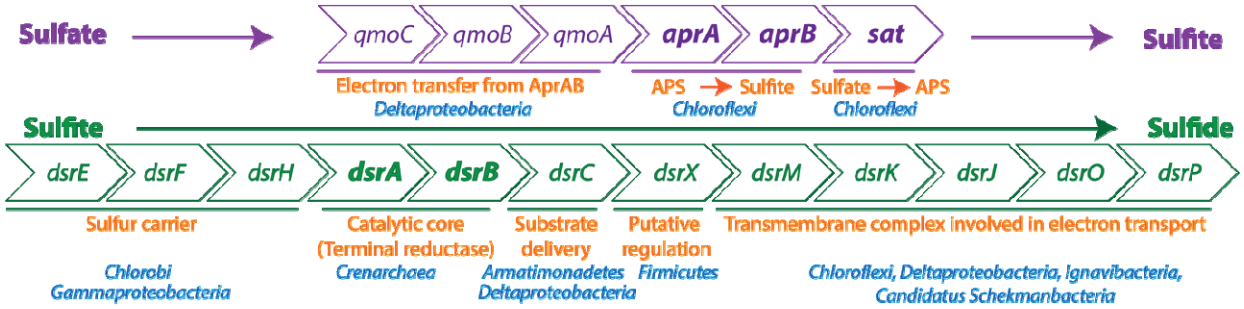

Fig. 3 Dsr operon structure and enzymatic roles of proteins involved in sulfate reduction in *Candidatus* Rokubacteria. Purple: genes involved in sulfate reduction to sulfite. Orange: putative enzymatic roles of genes, blue: microbial lineages with closest homologs as determined by phylogeny/blast against NCBI Genbank. APS refers to adenosine-5′-phosphosulfate. Green: genes involved in sulfite reduction to sulfide. This is the first case in which *dsrE*, *dsrF* and *dsrH* genes are present in organisms other than sulfur-oxidizing bacteria.

463 **Table 1. Details of lineages involved in dissimilatory sulfur cycling.**

Phylum	Number of Genomes reported in this study	Contribution to Sulfur Cycle	Mechanism	Source	Electron donor	
					Hydrogen	Organic C
Acidobacteria	3	Sulfate Reduction	dsr	A, D	Yes	Yes
Actinobacteria	3	Sulfate Reduction	dsr	-	Yes	Yes
Armatimonadetes	1	Sulfate Reduction	dsr	C	Yes	Yes
Candidatus Desantisbacteria	4	Sulfate Reduction	dsr	B, C	Yes	Yes
Candidatus Falkowbacteria	8	Putative sulfate reduction	unknown	A	Yes	Yes
Candidatus Hydrothermarchaeota	4	Sulfate Reduction	dsr	E	Yes	Yes
Candidatus Lambdaproteobacteria	5	Sulfite Reduction	dsr	A	Yes	Yes
Candidatus Muproteobacteria	14	Sulfur Oxidation	rdsr	A	Yes	Yes
Candidatus Omnitrophica	2	Sulfite Reduction	asr	A, B	Yes	Yes
Candidatus Riflebacteria	4	Sulfite Reduction	asr	A, B	Yes	Yes
Candidatus Rokubacteria	8	Sulfate Reduction	dsr	A	No	Yes
Candidatus Schekmanbacteria	1	Sulfate Reduction	dsr	A	No	Yes
Candidatus Zixibacteria	2	Sulfate Reduction	dsr	B	Yes	Yes
Chloroflexi	2	Sulfate Reduction	dsr	A	Yes	Yes
□-*proteobacteria*	34	Sulfate Reduction	dsr	A, C	Yes	Yes
Ignavibacteria	5	Sulfate Reduction	dsr	A, B	Yes	Yes
Nitrospinae	3	Sulfur Oxidation	rdsr	A	Yes	Yes
Nitrospirae	2	Sulfur Oxidation	rdsr	A	No	Yes
Nitrospirae	19	Sulfate Reduction	dsr	A	Yes	Yes
Planctomycetes	1	Sulfate Reduction	dsr	A	Yes	Yes

464

465 Sampling sources are indicated by letters: A - Aquifer at Rifle, Colorado, USA; B - Deep
466 subsurface in Japan; C - CO_2 geyser at Green River, Utah, USA; D - Glencore Mine, Canada; E -
467 Juan de Fuca ridge flank marine subsurface fluids. Newly identified lineages are shown in blue.