nature portfolio

Peer Review File

Evolutionary trajectory of microbial sulfur oxidation pathways recapitulates Earth's oxygenation history

Corresponding Author: Dr Song-Can Chen

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The study assembles a large, carefully curated bacterial dataset to trace the history of sulfur oxidation through the Sox system and place key innovations on an absolute timeline. The main picture is clear: the truncated Sox complex (SoxABXYZ) is ancient and largely vertically inherited; in-cluster SoxCD arises within Gammaproteobacteria and spreads to other phyla via transfers; coupling to rDsr occurs early but remains comparatively rare; and the recruitment of SoxCD aligns with or post-dates the Great Oxidation Event (GOE), with median ages ~2.1 Ga (uncorrelated) and ~2.4 Ga (autocorrelated) for the Gammaproteobacterial origin of the in-cluster module. The authors also document HGT into lineages such as Eremiobacterota and report that only ~2.7% of SoxB-hosting genomes carry both SoxCD and rDsr, consistent with metabolic flexibility under fluctuating oxygen. Together these results draw a coherent, geochemically plausible timeline for pathway assembly keyed to rising O_2.

Methods are, on the whole, appropriate and carefully executed. Module-wise gene trees for SoxAX, SoxBYZ, SoxCD, and rDsrAB are inferred on curated alignments; conflicts consistent with transfers (e.g., Eremiobacterota embedded within Pseudomonadota) are identified and discussed with ecological context. The species backbone is built from a concatenation of 27 ribosomal proteins plus 3 RNAP subunits and analyzed in IQ-TREE under LG+R10; alternative backbones reflecting recent consensus were tested with KH/SH/AU and found statistically indistinguishable, and this consensus-compatible topology was used for dating.

For molecular dating, the paper uses PhyloBayes with CAT20 and a birth–death prior, exploring both autocorrelated lognormal and uncorrelated gamma relaxed clocks. Two alternative root priors are considered (uniform 4.40–3.46 Ga and a gamma prior centered near 3.95 Ga), and absolute calibrations are introduced as a small set of "fossil evidence and strictly aerobic metabolisms" (supplementary figure/table cited by the authors). This framework produces the headline ages for SoxCD and related nodes and is presented as robust across the chosen clock and root-prior settings.

The main limitation is calibration handling. The species backbone uses a single site-homogeneous model (LG+R10) for the concatenation; while the authors test alternative backbones with likelihood topology tests, they do not show whether a site-heterogeneous model for the concatenation would materially shift deep splits that feed the clock. More importantly, although clock model and root-prior choices are varied, the absolute age calibrations themselves are not analyzed for sensitivity, and several are not expressed in standard form (node definition, stem vs. crown, minimum vs. maximum, and rationale). In Table S7, "Oxygenic Cyanobacteria > 2320—oxygenation of the atmosphere" is listed without saying whether the bound applies to the stem or crown Cyanobacteriia, and the text/figure does not disambiguate the placement. If this is a stem minimum, tying the onset of oxygenic photosynthesis to ≥ 2.32 Ga is conservative given evidence for pre-GOE oxygenic photosynthesis; if it is a crown minimum, the rationale is unclear because oxygenic photosynthesis evolved on the stem, and the LCA of extant oxygenic cyanobacteria could post-date the GOE. Similar issues apply to the GOE-linked maxima, the root interval (4.40–3.46 Ga), and other aerobic-metabolism constraints (Aerobic Nitrososphaerales < 2320; Aerobic Marinimicrobia < 2320; Nitrite oxidizers < 2320): the manuscript cites them but does not spell out node placement and justification in standard terms, and the robustness analysis does not vary these bounds. Given how strongly such constraints can steer absolute dates, this needs to be quantified.

Potential improvements:
1. Provide, for each calibration, a single paragraph, standard-form statement: node name and definition; stem or crown, e.g.

stem oxygenic cyanobacteria, minimum or maximum; numeric value(s); and the biological/geological rationale.
2. Add a calibration-sensitivity analysis that perturbs (i) the root minimum/maximum, in particular removing the minima and changing the maxima (ii) the GOE-linked maxima, and (iii) uses an appropriate oxygenic-cyanobacteria bound . Report how posterior medians and 95% CIs for the key innovation nodes shift and discuss how results are impacted.
3. Consider a site-heterogeneous alternative for the concatenated species backbone (or show that re-estimating the backbone under such a model leaves the dated nodes of interest stable), since deep bacterial splits can be model-sensitive.

This is a potentially impactful and thorough study with modern phylogenetic and dating practice. The results are biologically plausible and methodologically careful in most respects. To make the timescale fully persuasive, the manuscript should quantify how much the absolute age calibrations drive the estimates and state those calibrations in standard, reproducible terms. With that addition, the central claims would rest on a much firmer footing.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
In “Evolutionary trajectory of microbial sulfur oxidation pathways recapitulates Earth’s oxygenation history,” Ren et al. examine the evolutionary history of the Sox pathway involved in microbial sulfur oxidation, examining both its distribution across the tree of life (including viruses) as well as timing the emergence of key enzymes in deep time. Overall, many aspects of this study were carefully done—I commend the authors for their careful work in identifying homologs of the sox and dsr genes that most likely retain the sulfur oxidation or reduction function, and for comparing various clock models and comparing their results to previous molecular clocks. The results reported here support and provide interesting context on previous studies.

I have two main suggestions/concerns: first, the molecular clocks were generated with a quite narrow set of calibration points that are based almost entirely on the timing of the oxygenation of the atmosphere and could be expanded; second, much of the work in identifying the clade and timing for the origin of these genes involves comparing the topology of the gene trees with that of the species tree largely by eye. There are more robust pipelines for estimating the most likely clade in which various genes have arisen (such as ALERAX, for example) and I wonder if the results would be the same if these more robust approaches were applied.

More detailed comments are below.

Introduction
Line 77-81: it would be helpful to have a citation here for the functioning of the truncated variant without SoxCD: what studies have shown this in lab?

Methods

Phylogenetic tree reconstruction:

Lines 617: Building trees of sox genes: Last I checked, the automated model selection tool in IQ-TREE does not include complex mixture models, which have been shown to be more effective when resolving deep time phylogenies. This could be important for the results because tree topology does matter when doing phylogenetic reconciliation. It would be worth checking topologies generated using complex mixture models.

Lines 628 onward
Similarly, for the tree of life, the authors used the LG+R10 model. It is possible to use ModelFinder and apply complex mixture models to the tree of life as well; what was the rationale behind dictating the model in this fashion? Might instead recommend trying different aa substitution models for each marker protein using ModelFinder and allowing for complex mixture models in IQTREE, using a partitioned analysis with the -p option in IQ-TREE to allow each partition to have their own evolutionary rate. I can see that the authors also did some topological constraints (and grafting/pruning) after the fact, so the end result may be moot, but some rationale here is merited.

Calibration points: The calibration points used here effectively implies that certain aerobic clades couldn’t have emerged before the GOE. This seems likely, but as Davin et al (and others) have shown, there were likely some aerobic clades that emerged earlier (with accompanying “whiffs” of oxygen), so this isn’t completely 100% reliable. There are additional constraints that could be added here: for example, the origin of methanogens (last archaeal common ancestor, 4.4-3.46 based on Ueno 2006 and Cavalazzi et al 2021), total eukaryotes (placed at the MRCA of eukaryotic nuclear genomes and archaea) between 4.4 and 1.6 Ga (Pang 2013). There are later ones one could include too: purple sulfur bacteria at 1.631 Ga based on Brocks and Schaeffer 2008. I particularly suggest taking a look at the calibration points included in Davin et al (which I see is already cited in this paper), in particular Supplementary Methods section 4- the authors provide an extremely thorough and detailed explanation of their calibration points for a very similar tree of life.

Results
Figure 1c: I am a little confused about the “genome count” part of this figure. Is this the number of genomes included in the whole database, or was it the total number that contained a sox cluster? Please elaborate in the figure caption.

Figure 2:
A: What does “microbial lineages of ambiguous origin” mean in this context?
A: I wonder if, for the species tree, it would be easier to understand and read if this were a circular ToL showing which lineages have which subunits, or make it a supplementary figure. Then again, this may not be feasible visually.

Gene tree-species tree comparison: What is the rationale for deciding where the origin of the truncated sox system occurred, and the recruitment of dsr and soxCD? It seems that the authors used a version of parsimony by looking at the overall topology of the gene tree and species tree to determine the likely origin of the sox system, if I’m not mistaken. While this can yield some useful information, a more robust way of doing this would be to do a formal gene-tree and species-tree reconciliation using something like ALERAX or ecceTERA or something similar. This is important because this is also the basis for determining when some of these events occurred—the authors are dating a specific node in the tree where they assume these lineages first arose. Would the study yield the same answers for in which lineage these genes arose when if they were to do a formal reconciliation? I would probably recommend ALERAX for this (Morel et al., 2024) (https://doi.org/10.1093/bioinformatics/btae162) because it explicitly probabilities for gene births in different lineages.

I do think the results and interpretation overall so far make sense—it is consistent with previous studies, and the proposed oxidants are in line with what is observed in modern oceans (such as in OMZs), but it is important to ensure the methods are robust (i.e. calibration points and gene tree-species tree comparison).

Lines 359 onwards: For Random Forest model: would be nicer if there were a table showing all of the genomes and their phylogenetic/ecological markers, and just visualize this on a tree (this kind of feature is available in iToL, for example). For example, they could show the clades and their markers, which would allow the reader to look for patterns visually, accompanied by the analysis. This would likely be easier to understand than some of the plots shown in Figure 4.

Identification of viral proteins: please provide more information about how you distinguished lytic vs lysogenic viruses—did the authors search for specific genes (ie integrases), use a specific pipeline (VIBRANT), etc? I could not find this in the materials and methods section.

Figure 5b: I’m not convinced that the 90% RED dotted line is useful (it impedes the tree a bit), and the circles could be something a bit easier to interpret (bar charts?). Finally, why isn’t there a label for the taxonomic designation below Alphaproteobacteria and Gammaproteobacteria for the farthest right tree?

(Remarks on code availability)
I did not look at the code in great detail, but checked that it was available. I did not attempt to run it. This code seems to be specifically for identifying sox, rdsr, and shdr homologs in GTDB, which was one part of the bioinformatics analysis conducted fort his manuscript.

Version 1:

Reviewer comments:

Reviewer #2

(Remarks to the Author)
The authors have satisfactorily addressed all of my comments and concerns. The detailed explanation of the calibration points is particularly helpful, as added to the supplementary material. They appear to have added extra calibration points and new models of evolution and conducted extensive sensitivity testing to demonstrate that their results are robust to the evolutionary model and underlying assumptions, and they have conducted formalized gene tree-species tree reconciliation that shows their results are consistent with the parsimony analysis. I commend the authors for their nice work on this paper.

(Remarks on code availability)

Response to Reviewer's Comments

We are very happy that our study was generally well received by the reviewers and are grateful for their valuable comments, which helped us to improve the manuscript. We have addressed all comments in our **point-by-point response** and revised the manuscript accordingly. Please note that the numbering and order of figures and tables in the main text have been updated to accommodate the new material added during revision.

Our revision is based on extensive **molecular clock dating analyses** and **gene-species tree reconciliation** that, in brief, now clearly show that **our inferences regarding the origin and geological timing of Sox pathways are robust against alternative fossil calibrations and evolutionary models**.

Please find below a brief summary of the main analyses and outcomes.

Documentation of an expanded set of calibrations for Bayesian molecular dating analyses. Two additional fossil calibrations were adopted, and **standardized information was provided for each calibration point, including phylogenetic placement, maximum/minimum bounds, and the biological or geochemical justification**. Alternative hypotheses concerning the age of calibrated microbial groups and their incorporation into calibration sensitivity analyses were also described.

Consistent age estimates of Sox pathways under alternative calibration strategies. Multiple analytic scenarios that account for alternative combinations or bounds of calibrations were conducted under the Bayesian framework. **The resulting estimates indicate our original estimates concerning the geological ages of Sox pathways, relative to the timing of the Great Oxidation Event (GOE), remain robust.**

Phylogenetic reconstruction using site-heterogeneous substitution models yields congruent topologies. Complex mixture models and/or partition models were applied to reconstruct both gene and species trees. **All analyses recovered the same deep splits as in our original topologies, with key monophyletic clades for calibration and phylogenetic inference remaining stable.** Minor differences at the terminal branches do not impact divergence time estimate or phylogenetic reconciliation.

Gene-species tree reconciliation recapitulates our manual inference regarding the emergence of Sox pathway. Phylogenetic reconciliation on representative datasets **mapped the origin of Sox components to the same ancestral nodes in the species tree as initially proposed**. These results remain robust across evolutionary models of varying complexity used for tree reconstruction. Additionally, the geological timing of the stepwise assembly of the Sox system is further cross-validated via reconciliation analyses of previously available genome datasets.

Reviewer #1 (Remarks to the Author):

The study assembles a large, carefully curated bacterial dataset to trace the history of sulfur oxidation through the Sox system and place key innovations on an absolute timeline. The main picture is clear: the truncated Sox complex (SoxABXYZ) is ancient and largely vertically inherited; in-cluster SoxCD arises within Gammaproteobacteria and spreads to other phyla via transfers; coupling to rDsr occurs early but remains comparatively rare; and the recruitment of SoxCD aligns with or post-dates the Great Oxidation Event (GOE), with median ages ~2.1 Ga (uncorrelated) and ~2.4 Ga (autocorrelated) for the Gammaproteobacterial origin of the in-cluster module. The authors also document HGT into lineages such as Eremiobacterota and report that only ~2.7% of SoxB-hosting genomes carry both SoxCD and rDsr, consistent with metabolic flexibility under fluctuating oxygen. Together these results draw a coherent, geochemically plausible timeline for pathway assembly keyed to rising O_2.

Methods are, on the whole, appropriate and carefully executed. Module-wise gene trees for SoxAX, SoxBYZ, SoxCD, and rDsrAB are inferred on curated alignments; conflicts consistent with transfers (e.g., Eremiobacterota embedded within Pseudomonadota) are identified and discussed with ecological context. The species backbone is built from a concatenation of 27 ribosomal proteins plus 3 RNAP subunits and analyzed in IQ-TREE under LG+R10; alternative backbones reflecting recent consensus were tested with KH/SH/AU and found statistically indistinguishable, and this consensus-compatible topology was used for dating.

For molecular dating, the paper uses PhyloBayes with CAT20 and a birth–death prior, exploring both autocorrelated lognormal and uncorrelated gamma relaxed clocks. Two alternative root priors are considered (uniform 4.40–3.46 Ga and a gamma prior centered near 3.95 Ga), and absolute calibrations are introduced as a small set of “fossil evidence and strictly aerobic metabolisms” (supplementary figure/table cited by the authors). This framework produces the headline ages for SoxCD and related nodes and is presented as robust across the chosen clock and root-prior settings.

Response: We thank reviewer #1 for the careful and constructive review, as well as the overall positive evaluation of our manuscript.

The main limitation is calibration handling. The species backbone uses a single site-homogeneous model (LG+R10) for the concatenation; while the authors test alternative backbones with likelihood topology tests, they do not show whether a site-heterogeneous model for the concatenation would materially shift deep splits that feed the clock. More importantly, although clock model and root-prior choices are varied, the absolute age calibrations themselves are not analyzed for sensitivity, and several are not expressed in standard form (node definition, stem vs. crown, minimum vs. maximum, and rationale). In Table S7, “Oxygenic Cyanobacteria > 2320—oxygenation of the atmosphere” is listed without saying whether the bound applies to the stem or crown Cyanobacteriia, and the text/figure does not disambiguate the placement. If this is a stem minimum, tying the onset of oxygenic photosynthesis to ≥ 2.32 Ga is conservative given evidence for pre-GOE oxygenic photosynthesis; if it is a crown

minimum, the rationale is unclear because oxygenic photosynthesis evolved on the stem, and the LCA of extant oxygenic cyanobacteria could post-date the GOE. Similar issues apply to the GOE-linked maxima, the root interval (4.40–3.46 Ga), and other aerobic-metabolism constraints (Aerobic Nitrososphaerales < 2320; Aerobic Marinimicrobia < 2320; Nitrite oxidizers < 2320): the manuscript cites them but does not spell out node placement and justification in standard terms, and the robustness analysis does not vary these bounds. Given how strongly such constraints can steer absolute dates, this needs to be quantified.

Response: We agree with the reviewer's comments regarding the limitation of our calibration strategies. To address these issues, we have now (1) reconstructed alternative species trees using site-heterogeneous models (Fig. S3), (2) provided standardized, detailed description for each calibration point in the Methods section of Supplementary Information (SI, line 54-174), (3) performed extensive sensitivity analyses where we perturbed boundaries of calibration constraints or employed alternative species backbone topologies (Fig. S5 and Table S4). Specifically, we designed **eight** calibration scenarios to accommodate alternative hypotheses concerning the timing of root, total-group Cyanobacateriia, crown-group Cyanobacteriia, and GOE-associated microbial lineages. For instance, we relaxed our original constraint of crown-group Cyanobacteriia to allow its origin to post-date the GOE, and imposed constraints on the total-group oxygenic Cyanobacteriia to reflect the possibility that oxygenic photosynthesis arose along the stem of extant oxygenic Cyanobacteriia prior to the GOE. Additionally, we implemented **two** scenarios to test the impact of alternative backbone topologies reconstructed using complex mixture models and partition models, while maintaining our original calibration constraints. As detailed in the point-by-point responses below, our analyses support initial interpretation of our data.

Potential improvements:
1. Provide, for each calibration, a single paragraph, standard-form statement: node name and definition; stem or crown, e.g. stem oxygenic cyanobacteria, minimum or maximum; numeric value(s); and the biological/geological rationale.

Response: We have now provided a detailed description of all calibrations used in the molecular dating analyses (SI Methods line 54-174). We included seven calibration points in our initial analysis, and now extend this set to nine by incorporating total-group Cyanobacteriia and Chromatiaceae (Fig. S6 and Table S4), as proposed by Davín et al. (2025). For each calibration, we now specify the phylogenetic group, age boundaries, and justification in standardized format. Moreover, we have added a '*sensitivity analysis*' section detailing how each calibration constraint was varied across **eight** analytic scenarios to account for uncertainties in linking geological records to species phylogeny (**see the response below**).

2. Add a calibration-sensitivity analysis that perturbs (i) the root minimum/maximum, in particular removing the minima and changing the maxima (ii) the GOE-linked maxima, and (iii) uses an appropriate oxygenic-cyanobacteria bound. Report how posterior medians and 95% CIs for the key innovation nodes shift and discuss how results are impacted.

Response: We conducted extensive sensitivity analysis by varying constraints relative to our initial setup (*scenario 1–4*) across **eight** analytic scenarios (scenario *5–12*). These adjustments include (i) modifying root constraints by increasing the root maximum to 4.52 Ga (*scenario 5*), relaxing the root minimum to 3.225 Ga (*scenario 6*), or removing the root minimum (*scenario 7*), (ii) applying minimum bound to total-group Cyanobacteriia (*scenario 8* and *9*), relaxing minimum constraint on crown-group Cyanobacteriia to a post-GOE age (*scenario 8* and *9*), imposing the origin of the crown before the GOE (*scenario 10*), (iii) relaxing the GOE-linked maxima by allowing crown groups of aerobic lineages to evolve before the GOE (*scenario 10*), or removing entirely the GOE-linked maxima (*scenario 11*), and (iv) adding an additional calibration point for the total-group Chromatiaceae (*scenario 12*). The adjustments for each calibration are justified in SI Methods line 54-174. The yielded posterior distribution, medians and 95% CIs for the key innovation nodes are summarized in Fig. S5 and Table S4. Across all scenarios, the results are consistent with our initial estimates (Fig. 2d and Table S3), all placing the origins of truncated Sox system (①) and its association with rDsr before the GOE (③), and the in-cluster SoxCD (④) during or after the GOE. We therefore conclude our findings are robust against alternative calibration strategies.

Results and methods have been updated in the main text (line 189-190, 192, 262, and 326-328) and the Methods section (line 699-706), respectively.

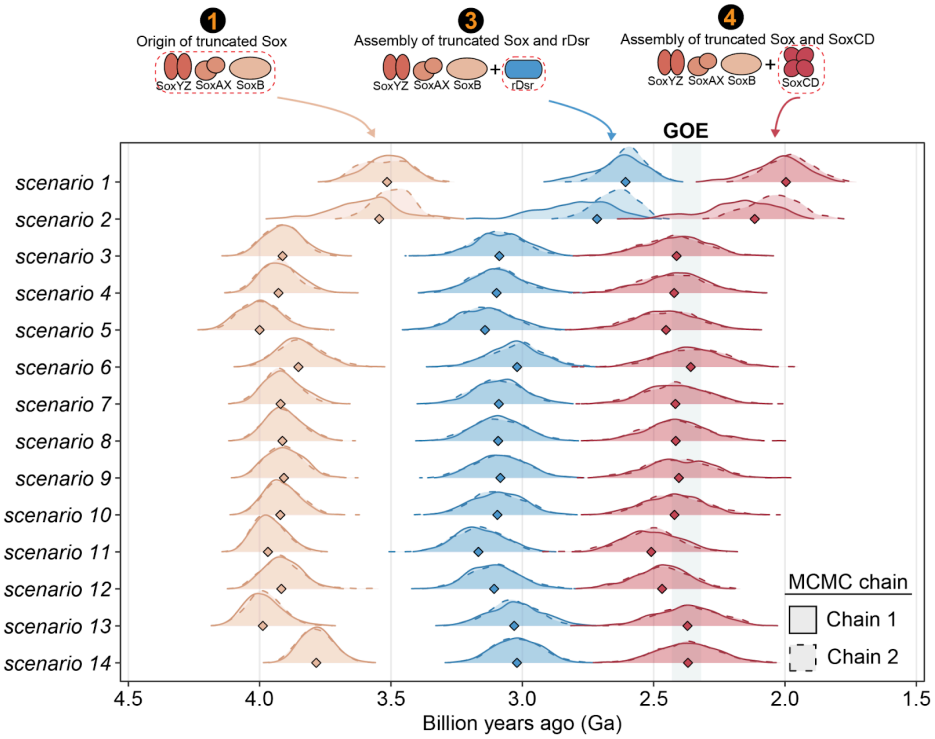

Figure S5 | Posterior probability density distributions for the ages of three key species nodes. Age distributions are shown for the nodes giving rise to SoxABXYZ (①), Sox-associated rDsr (③), and in-cluster SoxCD (④) under 14 alternative Bayesian analytic scenarios (Table S3-4). Diamonds indicate the medians. For each scenario, results from two independent Markov Chain Monte Carlo (MCMC) chains are displayed. Despite high dimensionality of the model, independent chains recover overall consistent posterior distributions for nodes ①, ②, and ④, providing robust timing for the assembly of Sox pathways relative to the Great Oxidation Event (GOE).

Table S4 | Sensitivity analyses of molecular dating with alternative temporal constraints and species backbones. The medians and 95% credible intervals for divergence times of nodes giving rise to SoxABXYZ, Sox-associated rDsr, and in-cluster SoxCD are shown across ten analytic scenarios. Uncorrelated molecular clock model and uniform root prior were applied to all analytic scenarios. Divergence time estimates and calibration points are expressed in billion years ago (Ga). The calibrations are derived from Martinez-Gutierrez et al. and Davín et al.

Adjustments[1]	Root age			Cyanobacteriia age		GOE-related maxima		Chromatiaceae age	Species backbones	
	Scenario 5	*Scenario 6*	*Scenario 7*	*Scenario 8*	*Scenario 9*	*Scenario 10*	*Scenario 11*	*Scenario 12*	*Scenario 13*	*Scenario 14*
Species tree[2]	Consensus	Consensus	Consensus	Consensus	Consensus	Consensus	Consensus	Consensus	Complex	Partition
Calibrations[3]										
LUCA (*C1*)	(4.52, 3.46)	(4.40, 3.225)	(4.40, -)	(4.40, 3.46)	(4.40, 3.46)	(4.40, 3.46)	(4.40, 3.46)	(4.40, 3.46)	(4.40, 3.46)	(4.40, 3.46)
LBCA (*C2*)	(4.52, 3.46)	(4.40, 3.225)	(4.40, 3.46)	(4.40, 3.46)	(4.40, 3.46)	(4.40, 3.46)	(4.40, 3.46)	(4.40, 3.46)	(4.40, 3.46)	(4.40, 3.46)
LACA (*C3*)	(4.52, 3.46)	(4.40, 3.225)	(4.40, 3.46)	(4.40, 3.46)	(4.40, 3.46)	(4.40, 3.46)	(4.40, 3.46)	(4.40, 3.46)	(4.40, 3.46)	(4.40, 3.46)
Total Cyanobacteriia (*C4*)	(-, -)	(-, -)	(-, -)	(-, 3.225)	(-, 2.78)	(-, -)	(-, 3.225)	(-, -)	(-, -)	(-, -)
Crown Cyanobacteriia (*C5*)	(-, 2.32)	(-, 2.32)	(-, 2.32)	(-, 2.013)	(-, 2.013)	(-, 2.52)	(-, 2.013)	(-, 2.32)	(-, 2.32)	(-, 2.32)
Crown AOA (*C6*)	(2.32, -)	(2.32, -)	(2.32, -)	(2.32, -)	(2.32, -)	(2.52, -)	(-, -)	(2.32, -)	(2.32, -)	(2.32, -)
Crown Marinimicrobia (*C7*)	(2.32, -)	(2.32, -)	(2.32, -)	(2.32, -)	(2.32, -)	(2.52, -)	(-, -)	(2.32, -)	(2.32, -)	(2.32, -)
Crown NOB (*C8*)	(2.32, -)	(2.32, -)	(2.32, -)	(2.32, -)	(2.32, -)	(2.52, -)	(-, -)	(2.32, -)	(2.32, -)	-
Total Chromatiaceae (*C9*)	(-, -)	(-, -)	(-, -)	(-, -)	(-, -)	(-, -)	(-, 1.631)	(-, 1.631)	(-, -)	(-, -)
Convergence of Bayesian analyses										
Maxdiff[4]	0.22	0.27	0.12	0.67	0.78	0.10	0.21	0.61	0.18	0.54
Minimum ESS[5]	62	247	114	102	61	53	127	135	200	80
Divergence time of nodes inferred to give rise to Sox components (Ga)[6]										
SoxABXYZ (①)	4.00	3.86	3.92	3.91	3.91	3.92	3.97	3.92	3.99	3.78
	4.14–3.85	4.01–3.69	4.02–3.76	4.03–3.76	4.03–3.76	4.03–3.80	4.06–3.83	4.04–3.75	4.10–3.83	3.90–3.66
Sox-associated rDsr (③)	3.14	3.03	3.08	3.09	3.08	3.09	3.17	3.11	3.03	3.02
	3.34–2.94	3.22–2.85	3.28–2.90	3.27–2.90	3.27–2.90	3.28–2.91	3.33–2.97	3.29–2.91	3.22–2.82	3.19–2.85
In-cluster SoxCD (④)	2.45	2.37	2.42	2.42	2.40	2.42	2.51	2.47	2.37	2.37
	2.70–2.23	2.59–2.15	2.64–2.19	2.64–2.18	2.65–2.18	2.64–2.22	2.69–2.29	2.69–2.29	2.61–2.16	2.59–2.14

[1] The setup of Bayesian runs was adjusted to reflect alternative assumptions on the timing of root, oxygenic Cyanobacteriia, and GOE-related maxima, and on species backbones. Additionally, the temporal constraint for the total group of Chromatiaceae was applied in the *scenario 11* and *12*.

[2] Species trees inferred using the site-homogenous model (LG+R10; 'Consensus', Fig. S3b), the complex mixture model (LG+C60+G; 'Complex', Fig. S3c), or the partition model ('Partition', Fig. S3d) were used for molecular clock analyses. Note that the calibration *node C8* is absent in the partition topology.

[3] The node id of calibration points (Fig. S6) are indicated in paratheses. The abbreviations are defined in the footnote of Table S3.

[4] Maxdiff, maximum discrepancy between two independent chains across 12 parameters. Despite the sub-optimal maxdiff in *scenario 8, 9, 12,* and *14*, it has minimal impacts on the age estimates for node ①, ②, and ④, as two independent chains yielded highly similar results (Fig. S5).

[5] ESS, effective sampling size.

[6] The divergence time is shown as median and 95% credible interval (CI). The 95% CI was derived from two independent chains for analytic scenarios showing high maxdiff or low ESS.

3. Consider a site-heterogeneous alternative for the concatenated species backbone (or show that re-estimating the backbone under such a model leaves the dated nodes of interest stable), since deep bacterial splits can be model-sensitive.

Response: We have now reconstructed alternative species trees using two different site-heterogeneous models: (1) a complex mixture model implemented employing profile mixture model LG+C60+G, and (2) a partition model where each gene block was assigned an independent substitution model (Fig. S3c,d). The deep branching patterns recovered under both models are highly consistent with our consensus topology (Fig. S3b), with all nodes associated with the calibrations and the origins of Sox components robustly resolved (Table S1). The only exception is the failure of the partition model to yield monophyly for crown aerobic nitrite-oxidizing bacteria (NOB) for calibration. To assess how alternative topologies affect our age estimation, we dated both species trees using our initial calibration set, excluding crown-group NOB from the partition-model backbone (*scenario 13* and *14*). As shown in Fig. S5 and Table S4, the estimated divergence time of the nodes giving rise to truncated Sox, Sox-linked rDsr, and in-cluster SoxCD remain consistent with our initial results (**see the response below**). We have now revised the manuscript accordingly at line 189-190 and 677-684.

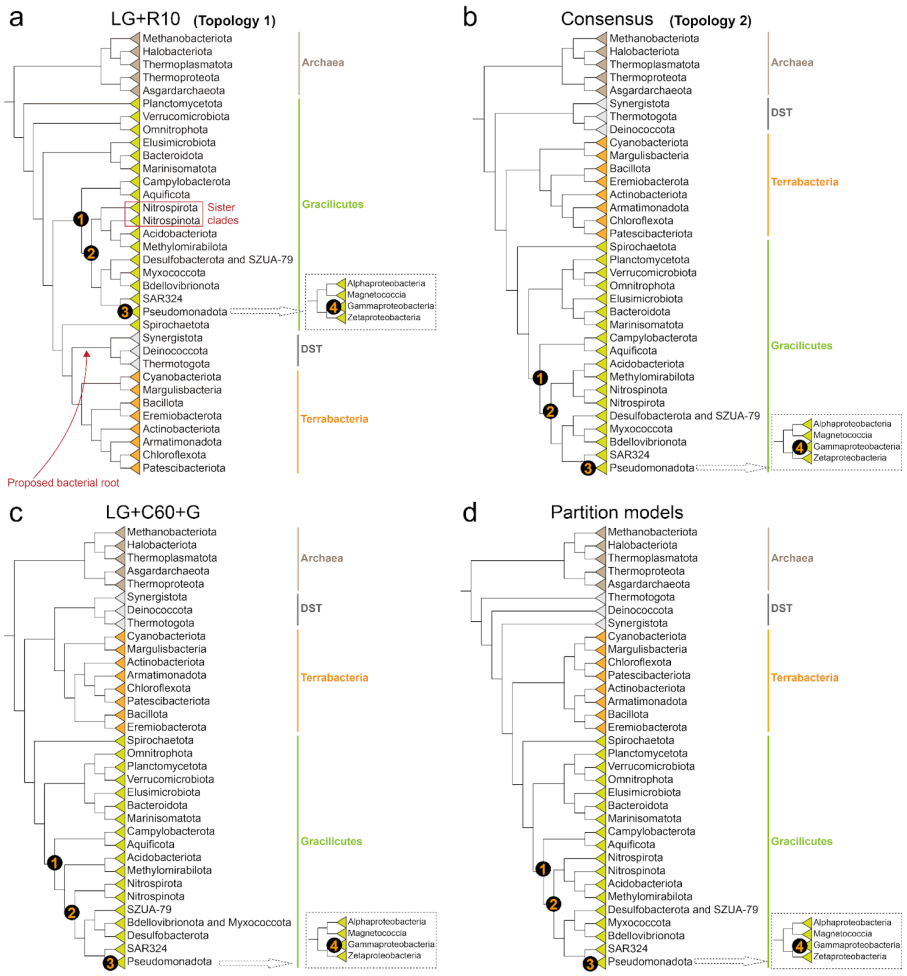

Figure S3 | Phylum-level branching patterns are overall congruent across species trees reconstructed using alternative evolutionary models in IQ-TREE. The trees were built using **a)** the basic substitution model LG+R10, **c)** the complex mixture model LG+C60+G; **d)**, a partition model that accounts for the heterogeneous substitution process of each gene block. For reference, a topology that reflects recent consensus of bacterial tree of life was pruned and regrafted from Topology 1 **(b)**. Enlarged subtrees (framed by dotted lines) show the branching patterns of four classes within the Pseudomonadota. Internal tree nodes inferred to give rise to Sox or rDsr systems are indicated by numbered circles (①, ②, ③, and ④); numbering corresponds to that that in Fig. 2a. DST is short for Deinococcota, Synergistota, and Thermotogota.

This is a potentially impactful and thorough study with modern phylogenetic and dating practice. The results are biologically plausible and methodologically careful in most respects. To make the timescale fully persuasive, the manuscript should quantify how much the absolute age calibrations drive the estimates and state those calibrations in standard, reproducible terms. With that addition, the central claims would rest on a much firmer footing.

Response: We thank the reviewer's constructive comments and encouraging assessments. In response, we have documented each calibration point in standard, reproducible terms, and showed the robustness of our age estimates to alternative use of fossil calibrations and tree topologies. We believe these results have strengthened the credibility of inferred timescale for the evolution of the Sox system.

Reviewer #2 (Remarks to the Author):

In "Evolutionary trajectory of microbial sulfur oxidation pathways recapitulates Earth's oxygenation history," Ren et al. examine the evolutionary history of the Sox pathway involved in microbial sulfur oxidation, examining both its distribution across the tree of life (including viruses) as well as timing the emergence of key enzymes in deep time. Overall, many aspects of this study were carefully done—I commend the authors for their careful work in identifying homologs of the sox and dsr genes that most likely retain the sulfur oxidation or reduction function, and for comparing various clock models and comparing their results to previous molecular clocks. The results reported here support and provide interesting context on previous studies.

Response: We gratefully thank the reviewer for the detailed reading of our manuscript, and the positive assessment of our work.

I have two main suggestions/concerns: first, the molecular clocks were generated with a quite narrow set of calibration points that are based almost entirely on the timing of the oxygenation of the atmosphere and could be expanded; second, much of the work in identifying the clade and timing for the origin of these genes involves comparing the topology of the gene trees with that of the species tree largely by eye. There are more robust pipelines for estimating the most likely clade in which various genes have arisen (such as ALERAX, for example) and I wonder if the results would be the same if these more robust approaches were applied.

Response: We thank the reviewer for these helpful suggestions and concerns. We have expanded the set of calibration points beyond those tied to the Great Oxidation Event (GOE), and performed extensive sensitivity analysis to demonstrate that our inferred time is robust to variations in calibration bounds (**see our response to reviewer #1**). Moreover, we have applied formal gene-species tree reconciliation analysis on representative datasets using Ranger-DTL or ecceTERA. These approaches yielded consistent results with our manual parsimony-based method regarding the emergence of different Sox components.

More detailed comments are below.

Introduction

Line 77-81: it would be helpful to have a citation here for the functioning of the truncated variant without SoxCD: what studies have shown this in lab?

Response: The physiological role of the truncated Sox system has been demonstrated in model sulfur-oxidizing microorganisms, including *Allochromatium vinosum* (Hensen et al., 2006), *Chlorobaculum tepidum* (Holkenbrink et al., 2011), and *Aquifex aeolicus* (Boughanemi et al., 2016). We have now cited these studies at line 79-81 in the revised manuscript.

Methods

Phylogenetic tree reconstruction:

Lines 617: Building trees of sox genes: Last I checked, the automated model selection tool in IQ-TREE does not include complex mixture models, which have been shown to be more effective when resolving deep time phylogenies. This could be important for the results because tree topology does matter when doing phylogenetic reconciliation. It would be worth checking topologies generated using complex mixture models.

Response: We have now applied complex mixture models to reconstruct the gene phylogeny of SoxBYZ, SoxAX, SoxCD, rDsrAB, and sHdr. For each dataset, we chose the optimal mixture model by extending the IQ-TREE model selection procedure to include profile mixture models C10 to C60 (Table S9). The resulting phylogenies recovered identical deep branching patterns as previously observed, i.e., (1) a clear division between Clade I and II, including the monophyletic groups thereof, in the phylogenies of SoxBYZ and SoxAX; (2) separation of Burkholderiales and other Gammaproteobacteria orders in the SoxCD phylogeny; (3) a split between Nitrospirota DsrAB and Pseudomonadota DsrAB; and (4) three major clades corresponding to SZUA-79, Aquificota, and Pseudomonadota in the sHdr phylogeny (Fig. S4). These results demonstrate that the gene tree topology remains stable even under increased model complexity. Consistent with this, phylogenetic reconciliation analyses confirmed minor impacts of alternative evolutionary models for gene tree reconstruction on our inferences regarding the origin of gene components (**see the response below**).

We have updated these results in SI Results (line 233-254), and added tree reconstruction procedures using complex mixture models in main text (line 637-643).

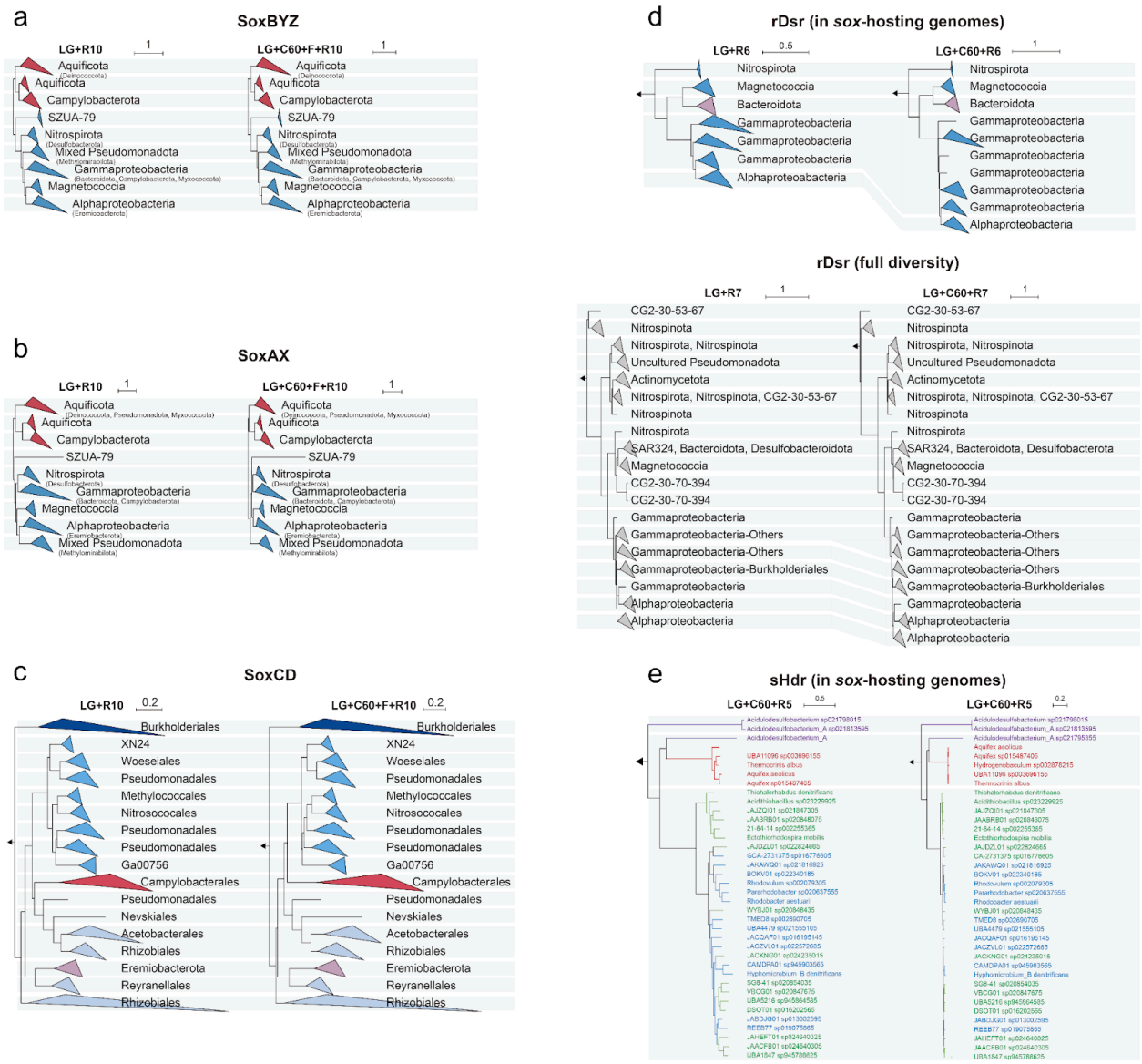

Figure S4 | The overall phylogenetic topology is robust to the choices of evolutionary models used to reconstruct trees for SoxBYZ (a), SoxAX (b), SoxCD (c), rDsrAB (d), and sHdrAB1C1B2C2 (e). In each panel, trees reconstructed with (right) or without (left) considering the complex mixture models are shown. The optimal substitution model selected by IQ-TREE ModelFinder is indicated above the tree. Identical bipartitions of two compared trees are indicated by gray shade. The coloring scheme of the clades is consistent with that in Fig. 2b, Fig. 2c, Fig. 3b, Fig. S9, and Fig. S10.

Lines 628 onward

Similarly, for the tree of life, the authors used the LG+R10 model. It is possible to use ModelFinder and apply complex mixture models to the tree of life as well; what was the rationale behind dictating the model in this fashion? Might instead recommend trying different aa substitution models for each marker protein using ModelFinder and allowing for complex mixture

models in IQTREE, using a partitioned analysis with the -p option in IQ-TREE to allow each partition to have their own evolutionary rate. I can see that the authors also did some topological constraints (and grafting/pruning) after the fact, so the end result may be moot, but some rationale here is merited.

Response: We selected LG+R10 following the work by Martinez-Gutierrez et al. (2021), who identified this model as optimal for describing the evolutionary process of our protein set (27 ribosomal proteins and 3 RNA polymerases). We have now reaffirmed this by applying the standard ModelFinder algorithm to our dataset (Table S9). Moreover, we re-calculated the species tree using the optimal mixture model (LG+C60+G; Table S9) and the partition analyses (Table S11). Both approaches produced highly consistent clustering patterns for the major archaeal and bacterial phyla compared to our reference tree (Fig. S3). All three trees recover Terrabacteria and Gracilicutes as two monophyletic bacterial groups, and yield congruent internal nodes that were inferred to give rise to Sox or rDsr components. The main topological difference lies in the placement of Deinoccota, Synergistota, and Thermotogota (DST). Under the basic and complex substitution models, DST forms a monophyletic group, whereas these taxa branch as three independent basal bacterial lineages in the partitioned analysis. Additionally, under the partitioned models, monophyly of the crown aerobic nitrite-oxidizing bacteria (Nitrospirota and Nitrospinota, NOB) is not recovered. Nevertheless, these topological differences do not affect our main conclusions based on phylogenetic reconciliation (Table S1) or molecular dating (Fig. S5 and Table S4) (**see the response above**).

We have added the species tree reconstruction procedures using site-heterogenous models in the main text at line 677-684.

Calibration points: The calibration points used here effectively implies that certain aerobic clades couldn't have emerged before the GOE. This seems likely, but as Davin et al (and others) have shown, there were likely some aerobic clades that emerged earlier (with accompanying "whiffs" of oxygen), so this isn't completely 100% reliable. There are additional constraints that could be added here: for example, the origin of methanogens (last archaeal common ancestor, 4.4-3.46 based on Ueno 2006 and Cavalazzi et al 2021), total eukaryotes (placed at the MRCA of eukaryotic nuclear genomes and archaea) between 4.4 and 1.6 Ga (Pang 2013). There are later ones one could include too: purple sulfur bacteria at 1.631 Ga based on Brocks and Schaeffer 2008. I particularly suggest taking a look at the calibration points included in Davin et al (which I see is already cited in this paper), in particular Supplementary Methods section 4- the authors provide an extremely thorough and detailed explanation of their calibration points for a very similar tree of life.

Response: We thank the reviewer for recommending Davin's paper and for thoughtful suggestions regarding alternative and additional calibration points. Following this recommendation, we have expanded our initial calibration set (including one for methanogenic archaea) by adding constraints on total-group Cyanobacteriia and Chromatiaceae (purple sulfur bacteria) (see *node C4* and *node C9* in Fig. S6). We have to leave out many Eukaryotes calibrations used in Davin's paper, because our species tree lacks representatives of

Eukaryotes and their chloroplastids and/or mitochondria. For each calibration point, we have provided a detailed, standardized description, and conducted sensitivity analysis by systematically disturbing its bounds across **eight** analytic scenarios. To show an example, we relaxed GOE-linked maxima to allow early emergence of aerobic clades in the late Archean period (~2.5 Ga), prior to the GOE, during intervals when 'oxygen whiffs' have been documented (see *scenario 10* and SI Methods). The divergence times estimated under these alternative scenarios remain consistent with our original results (Fig. S5; Table S4), reinforcing the robustness of our inferences regarding the evolutionary timing of the Sox system (**see the response above**).

Results

Figure 1c: I am a little confused about the "genome count" part of this figure. Is this the number of genomes included in the whole database, or was it the total number that contained a sox cluster? Please elaborate in the figure caption.

Response: The "genome count" indicates the number of genomes containing *sox* gene cluster(s). We have revised the caption of Fig. 1c accordingly (line 160-161).

Figure 2:

A: What does "microbial lineages of ambiguous origin" mean in this context?

Response: We meant the microbial lineages that encode the Sox system of either vertical or horizontal origin. We have now specified this information in the legend of Fig. 2 to avoid ambiguity (line 224).

A: I wonder if, for the species tree, it would be easier to understand and read if this were a circular ToL showing which lineages have which subunits, or make it a supplementary figure. Then again, this may not be feasible visually.

Response: We have now shown the distribution of Sox and rDsr subunits across a circular ToL in a supplementary figure (Fig. S17), yet the pattern is not fully conspicuous due to the large tree size. Alternatively, we indicate the distribution and inheritance patterns of Sox and rDsr subunits directly in Fig. 2a at phylum level. We used circles of different shades to distinguish the presence and absence of subunits in each lineage, with the former further categorized by their inferred origin (lateral or vertical). We believe this provides better visualization of the distribution patterns across lineages.

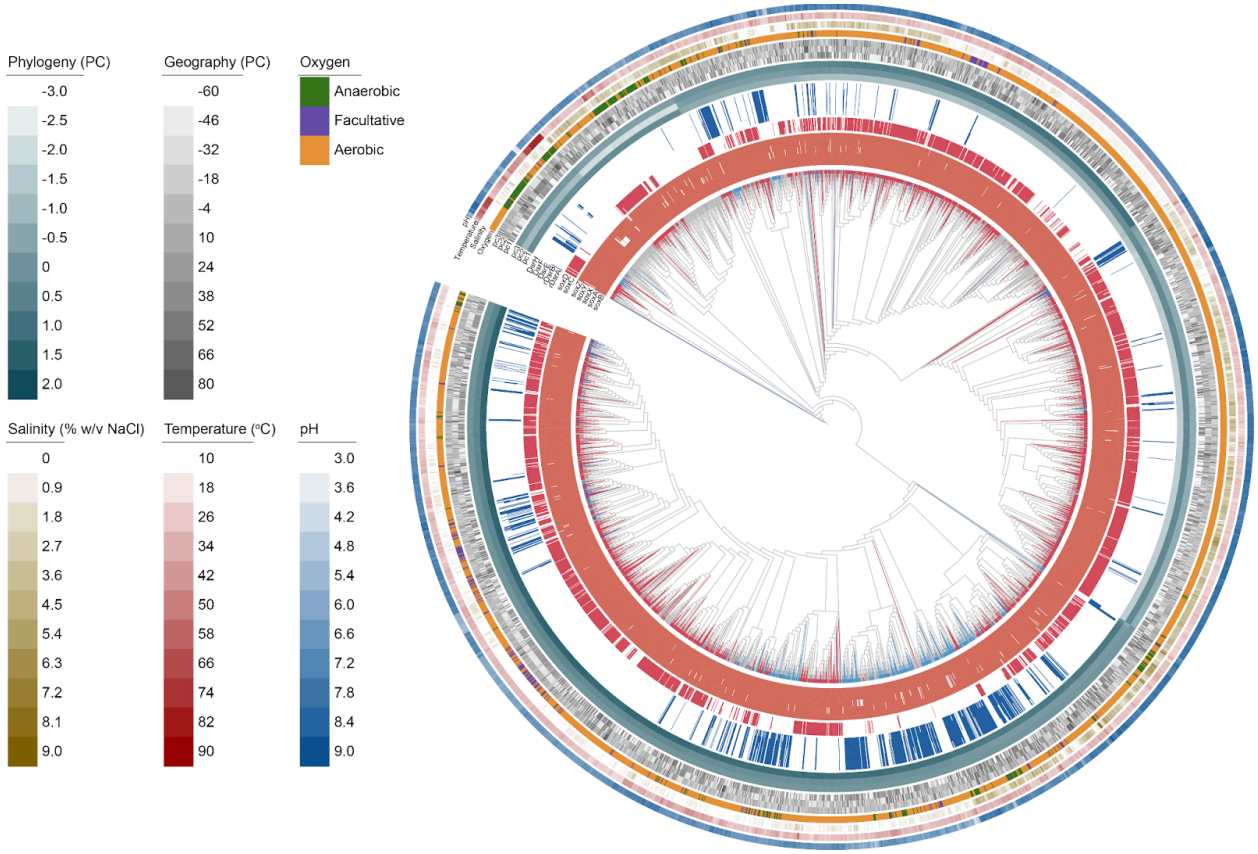

Figure S17 | The phylogenetic, geographical, and ecological traits of *sox*-hosting microorganisms. *Sox*-hosting microorganisms were classified into four categories based on the genomic presence or absence of key genes: (1) complete Sox system without rDsr (red branches), (2) truncated Sox system coupled with rDsr (blue branches), (3) complete Sox system coupled with rDsr (purple branches); and (4) truncated Sox system without rDsr (grey branches). The species tree was pruned from Tree of Life (GTDB r214) based on the representative *sox*-hosting species (n = 3,321, 87.5% of all) with complete metadata of the phylogenetic, geographic, and ecological traits. The inner rings (1–12) show the presence and absence of specific Sox or Dsr components. The middle rings (13–18) denote the first three principal components (PCs) decomposed based on phylogenetic or geographical distance. The outermost rings (19–22) indicate optimal growth conditions, including oxygen requirement, salinity, temperature, and pH. Phylogenetic inertia of alternative Sox pathways is evident, with lineages sharing the Sox pathways tending to cluster together. In addition, a clear association between oxygen preference and Sox pathways exists, with systems comprising truncated Sox and rDsr being over-represented among anaerobic lineages.

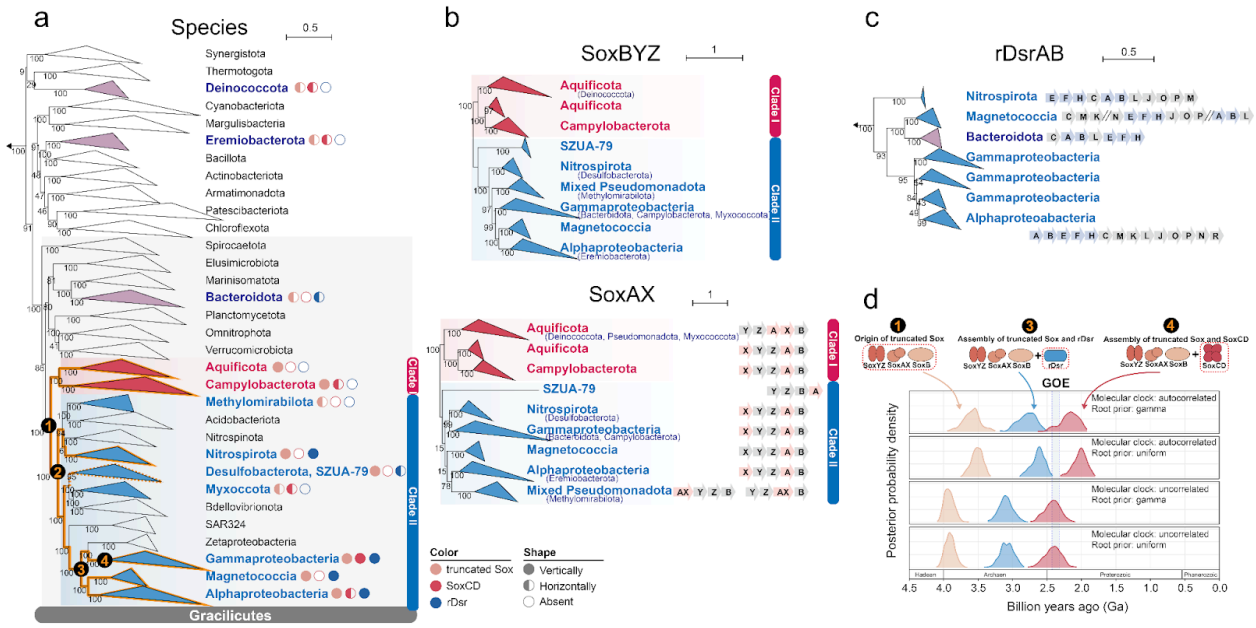

Figure 2 | Phylogenies of bacteria, SoxBYZ, SoxAX, and rDsrAB. (a) Bacteria species tree constructed through maximum likelihood estimation based on a concatenated alignment of 30 RNA polymerase subunits and ribosomal proteins as previously reported [62]. The root is placed according to Martinez-Gutierrez et al., Martinez-Gutierrez and Alyward, and Coleman et al. [45,62,63]. Circles beside species names show the distribution and inferred origin (vertical or lateral) of Sox and rDsr subunits across lineages. Vertical inheritance of the truncated Sox system is highlighted as orange branches. Microbial lineages (Desulfobacterota and SZUA-79) that encode Sox systems of either vertical or horizontal origin (Supplementary Results) are denoted with dashed lines. Major *sox*-encoding bacterial lineages are colored in red (Clade I), blue (Clade II), and purple (outside Clades I and II), respectively. Ancestral nodes corresponding to the origin of the truncated Sox system (①), the recruitment of rDsr (② or ③), and the recruitment of SoxCD (④) within Gammaproteobacteria, are indicated with orange numbers. **(b-c)** The maximum likelihood phylogeny of SoxBYZ (model LG+R10), SoxAX (model LG+R10), and rDsrAB (model LG+R6). Lineage colors correspond to panel (a). Taxa comprising a minority of the clade members or nested within the collapsed clades are shown with smaller font in parenthesis. The trees are collapsed based on the phylum-, class- or order-level classification of the host microorganisms. The uncollapsed trees can be found in Fig. S7 and Fig. S9. Typical operon structures of SoxABXYZ and rDsr for lineages are illustrated on the right side of their respective protein phylogenies. Values below the branches represent ultrafast bootstraps calculated by IQtree. The overall phylogenetic topology is robust to the choices of evolutionary models used to reconstruct species and gene trees (Fig. S3 and Fig. S4). **(d)** Posterior probability density for the age of three key nodes (①, ③, and ④) in panel (a), which represent key evolutionary events over natural history of the Sox system. The results yielded by four analytic settings (*scenario 1–4*, Table S3) considering different combinations of the molecular clock models (autocorrelated log normal *vs.* uncorrelated gamma) and root prior distribution (gamma vs. uniform) are shown in four subplots, with model parameters specified in top right.

Gene tree-species tree comparison: What is the rationale for deciding where the origin of the truncated sox system occurred, and the recruitment of dsr and soxCD? It seems that the authors used a version of parsimony by looking at the overall topology of the gene tree and species tree to determine the likely origin of the sox system, if I'm not mistaken. While this can yield some useful information, a more robust way of doing this would be to do a formal gene-tree and species-tree reconciliation using something like ALERAX or eccTERA or something similar. This is important because this is also the basis for determining when some of these events occurred—the authors are dating a specific node in the tree where they assume

these lineages first arose. Would the study yield the same answers for in which lineage these genes arose when if they were to do a formal reconciliation? I would probably recommend ALERAX for this (Morel et al., 2024) (https://doi.org/10.1093/bioinformatics/btae162) because it explicitly probabilities for gene births in different lineages.

Response: Yes, we determined evolutionary history of the Sox-related genes through manual comparison of gene tree with the species tree, guided by the principle of parsimony. We did not apply automated phylogenetic reconciliation on our full dataset (> 80,000 species with mean gene size > 3000), because direct reconciliation at this scale is computationally infeasible with current methods, which typically accommodate at most a few thousand taxa.

To enable formal reconciliation, we have now compiled a taxonomically balanced subsample of 569 representative genomes following a sampling strategy by Martinez-Gutierrez et al. (2021, 2023). On this reduced dataset, we reconciled gene/species tree using Ranger-DTL according to an established workflow (Nishihara et al., 2024). To ensure the robustness of results, we explored a range of biologically plausible costs for duplication, transfer, and loss events, and aggregated 100 optimal reconciliations randomly sampled from each cost combination. The present probability (PP) of gene birth along a specified lineage was then calculated as frequency across all aggregated samples.

The reconciliation procedure recapitulates our manual inferences: the earliest evolutionary event mapped to (1) the last common ancestor of Clade I and II for both SoxBYZ and SoxAX, (2) the crown Pseudomonadota for sox-associated rDsrAB, and (3) an ancient organism within Gammaproteobacteria clade for in-cluster SoxCD – all with high support (PPs > 0.90; Table S1). These results were robust against substitution models used for gene tree (basic vs. complex mixture models) and species tree inferences (basic, complex mixture, or partition models).

For independent cross-validation, we re-analyzed the data by Martoes et al. (2023) using our conservative homolog identification pipeline, followed by reconciliation with ecceTERA using exactly the same parameters and protocols of the original study. This confirmed the pre-GOE origin of SoxBYZ, SoxAX, and rDsrAB, and post-GOE origin of SoxCD (Table S2). Minor discrepancy with the original Marteos et al. (2023) analysis on the dates of DsrAB and SoxCD arises from (1) our deliberate focus on subclades that are directly associated with the truncated Sox system, and (2) the strictness of our homolog identification pipeline.

In summary, reconciliation analyses corroborate our conclusions on the origin of the truncated Sox system, and the recruitment of rDsr and SoxCD. We emphasize the value of large scale manual phylogenetic interpretation, which captures far greater microbial diversity and potentially alleviates sampling bias introduced by the subsampling required for computational tractability in automated reconciliation. This traditional approach remains essential and should be employed alongside the automatic methods to ensure robust and comprehensive evolutionary inferences.

We have updated these results in the SI (line 255-275), and added the reconciliation pipeline in the method section of main text (line 651-659) and SI (line 26-53).

Table S1 | The earliest evolutionary event of Sox or rDsr components inferred by gene-species tree reconciliation. For each gene set, phylogenetic reconciliation was conducted under six analytic scenarios, employing gene trees or species trees reconstructed using different substitution models. Results from each scenario were compiled from a wide range of cost values for duplication (C_D), transfer (C_T), and loss (C_L) events.

Gene set	C_D:C_T:C_L[1]	Gene tree Best-fit model[2]	Species tree Model[3]	Gene birth node[4]	Comments
soxBYZ	C_D:C_T=2:3, 1:1, 4:3, 5:3	LG+R6	LG+R10	① (S, PP = 1)	The earliest evolutionary event was inferred as a speciation event occurring at the last common ancestor (LCA) of clade I and II microbes (node ① in the species tree in Fig. 2 and S3) for both SoxBYZ and SoxAX. This estimation has high statistical support (present probability (PP) = 1) over a range of costs (C_D:C_T:C_L), and is robust against model complexity for tree reconstruction.
			LG+C60+G	① (S, PP = 1)	
			Partition	① (S, PP = 1)	
	C_L:C_T=1:6, 1:8, 1:12	LG+C40+F+R6	LG+R10	① (S, PP = 1)	
			LG+C60+G	① (S, PP = 1)	
			Partition	① (S, PP = 1)	
soxAX	C_D:C_T=2:3, 1:1, 4:3, 5:3	Q.pfam+I+G4	LG+R10	① (S, PP = 1)	
			LG+C60+G	① (S, PP = 1)	
			Partition	① (S, PP = 1)	
	C_L:C_T=1:6, 1:8, 1:12	LG4X+I+G	LG+R10	① (S, PP = 1)	
			LG+C60+G	① (S, PP = 1)	
			Partition	① (S, PP = 1)	
rDsrAB	C_D:C_T=2:3, 1:1, 4:3, 5:3	LG+R5	LG+R10	③ (T, PP = 1)	The origin of rDsrAB from *sox*-hosting microorganisms was inferred as a transfer event involving the LCA of Pseudomonadota (node ③). This inference is robust over a range of costs (C_D:C_T:C_L) and varying substitution models for tree reconstruction.
			LG+C60+G	③ (T, PP = 1)	
			Partition	③ (T, PP = 1)	
	C_L:C_T=1:6, 1:8, 1:12	LG+C60+F+R4	LG+R10	③ (T, PP = 1)	
			LG+C60+G	③ (T, PP = 1)	
			Partition	③ (T, PP = 1)	
soxCD	C_D:C_T=2:3, 1:1, 4:3, 5:3	Q.pfam+I+G4	LG+R10	④' (S, PP = 1)	SoxCD was predicted to originate at a descendent lineage of the LCA of the Gammaproteobacteria (node ④), which was denoted as node ④'. This is consistent with our estimation that recruitment of SoxCD was a recent event over evolutionary history of the Sox system.
			LG+C60+G	④' (S, PP = 1)	
			Partition	④' (S, PP = 1)	
	C_L:C_T=1:6, 1:8, 1:12	LG4X+I+G	LG+R10	④' (S, PP = 1)	
			LG+C60+G	④' (S, PP = 1)	
			Partition	④' (S, PP = 1)	

[1] All combinations (n = 12) of cost ratio C_D:C_T (2:3, 1:1, 4:3, 5:3) and C_L:C_T (1:6, 1:8, 1:12) were used for reconciliation. Results were compiled from 100 randomly sampled optimal reconciliations from each of 12 cost combinations.

[2] The best-fit model for each gene set was selected by ModelFinder with or without considering mixture models (e.g., C10-C60 and LG4X). Note that the optimal model for gene reconstruction is different from those in Table S9, due to different scales of dataset. The genes in this table were retrieved from 569 representative genomes, whereas the dataset shown in Table S9 was derived from > 80000 GTDB genomes.

[3] Species tree was reconstructed using basic substitution model (LG+R10), complex mixture (LG+C60+G), or partition models. Both basic and complex models were selected by ModelFinder in IQ-TREE.

[4] The column shows the node in the species tree that was predicted to give rise to the gene(s). The nodes ①, ③, and ④ in the species tree are shown in Fig. 2a and S3, which correspond to the last common ancestor (LCA) of clade I and II microorganisms, LCA of Pseudomonadota, and LCA of Gammaproteobacteria, respectively. Node ④' indicates an internal node within the Gammaproteobacterial clade. The earliest evolutionary event (S, speciation; T: horizontal gene transfer) and its presence probability (PP) are shown in paratheses.

Table S2 | The gene birth date (billion years ago; Ga) of SoxBYZ, SoxAX, rDsrAB, and SoxCD. The results were yielded by re-analyzing the genomic dataset from a previous study by Mateos et al. [10] The gene birth date was inferred as the earliest evolutionary event based on reconciliation of gene tree against a dated species tree using ecceTERA.

Gene set[1]	Gene tree Best-fit model[2]	Date or date range for the earliest gene event[3]	Comments
soxBYZ	LG+I+G4	2.59 Ga (S, PP = 1)	The earliest evolutionary event of SoxBYZ and SoxAX was inferred as a speciation event at 2.59 Ga, predating the GOE (~2.35 Ga). This timing is consistent with both our estimate and that of the original study, and is robust against model complexity for gene tree reconstruction.
	C50+I+G	2.59 Ga (S, PP = 1)	
soxAX	WAG+I+G4	2.59 Ga (S, PP = 1)	
	LG4M+R3	2.59 Ga (S, PP = 1)	
rdsrAB	LG+R5	2.86–2.50 Ga 2.68 Ga (T, PP = 1)	The timing of rDsrAB predates the GOE, which is consistent with our estimates but younger than the original study (3.3 Ga). This is likely due to our focus on the Sox-associated DsrA/B sequences, which constitute a subclade of the original DsrA/B tree.
	C60+R5	2.86–2.50 Ga 2.68 Ga (T, PP = 1)	
soxCD	Q.pfam+I+G4	0.90–0 Ga 0.45 Ga (D, PP = 0.63)	The timing of the in-cluster SoxCD postdates the GOE, which is consistent with our estimates but younger than the original study (2.77 Ga). This is likely because we only included the in-cluster copies in our analyses, whereas Mateos et al. treated paralogous SoxC, in-cluster SoxC, and potentially SorA as SoxC. Notably, the duplication event is mapped to a terminal tip in the species tree, and therefore the lower temporal boundary is effectively constrained to 0.
	C60+I+G	0.90–0 Ga 0.45 Ga (D, PP = 1)	

[1] *rdsrAB*: oxidative-type *dsrAB* that are associated with the Sox system; *soxCD*: *soxCD* that are nested within the *sox* gene cluster (in-cluster *soxCD*). Fewer sequences than Mateos et al. (2023) were retrieved for each gene set, because a more stringent homologs search pipeline was applied.

[2] The best-fit model for each gene set was selected by ModelFinder with or without considering mixture models (e.g., C10-C60 and LG4X). Note that the optimal model for gene reconstruction is different from those in Table S9, due to different scales of dataset. The genes in this table were retrieved from 871 representative genomes of Mateos et al. (2023), whereas the dataset shown in Table S9 was derived from > 80,000 GTDB genomes.

[3] The column shows the estimated age of the earliest evolutionary event for each gene set, as identified by ecceTERA. The earliest evolutionary event (S, speciation; T: horizontal gene transfer; D: duplication) and its presence probability (PP) are shown in parentheses. Speciation events are assigned to specific nodes with precise dates. Horizontal transfer and duplication events occur along branches, which are assigned to a date range spanning the ages of the mapped node and its parent. For horizontal gene transfers, the date range is further refined as the intersection of the recipient's and donor's branch date ranges. The midpoint of the date range is also reported.

I do think the results and interpretation overall so far make sense—it is consistent with previous studies, and the proposed oxidants are in line with what is observed in modern oceans (such as in OMZs), but it is important to ensure the methods are robust (i.e. calibration points and gene tree-species tree comparison).

Response: We have made every effort to ensure methodological robustness, and we believe the current analyses are sufficient to support our conclusions.

Lines 359 onwards: For Random Forest model: would be nicer if there were a table showing all of the genomes and their phylogenetic/ecological markers, and just visualize this on a tree (this kind of feature is available in iToL, for example). For example, they could show the clades and their markers, which would allow the reader to look for patterns visually, accompanied by the analysis. This would likely be easier to understand than some of the plots shown in Figure 4.

Response: We have compiled a comprehensive table listing all genomes together with their phylogenetic and ecological markers in Random Forest analysis (Supplementary Data 6). As the reviewer suggested, we have now added a new figure (Fig. S17; **see the response above**) that overlays these data onto the species tree to facilitate the visualization of the patterns.

Identification of viral proteins: please provide more information about how you distinguished lytic vs lysogenic viruses—did the authors search for specific genes (ie integrases), use a specific pipeline (VIBRANT), etc? I could not find this in the materials and methods section.

Response: We used VIBRANT v1.2.0 to distinguish lytic *vs.* lysogenic viruses using the command, which provide the annotation of virus lifestyles. We have added a relevant statement at line 783.

Figure 5b: I'm not convinced that the 90% RED dotted line is useful (it impedes the tree a bit), and the circles could be something a bit easier to interpret (bar charts?). Finally, why isn't there a label for the taxonomic designation below Alphaproteobacteria and Gammaproteobacteria for the farthest right tree?

Response: We used 90% RED dotted lines to highlight the recent divergence of viral Sox genes from their homologs in bacterial hosts. To increase the readability, we replaced the lines with light gray shading, and added a scale to clarify the circle sizes corresponding to the number of placed sequences. The bottom branches were initially left unlabelled because they represent paralogous SoxCD proteins whose physiological roles remain unknown. Viral sequences placed within these branches were not regarded as genuine auxiliary genes mediating Sox-based metabolism. To avoid confusion, we have labelled the paralogous branch, and removed the associated circles.

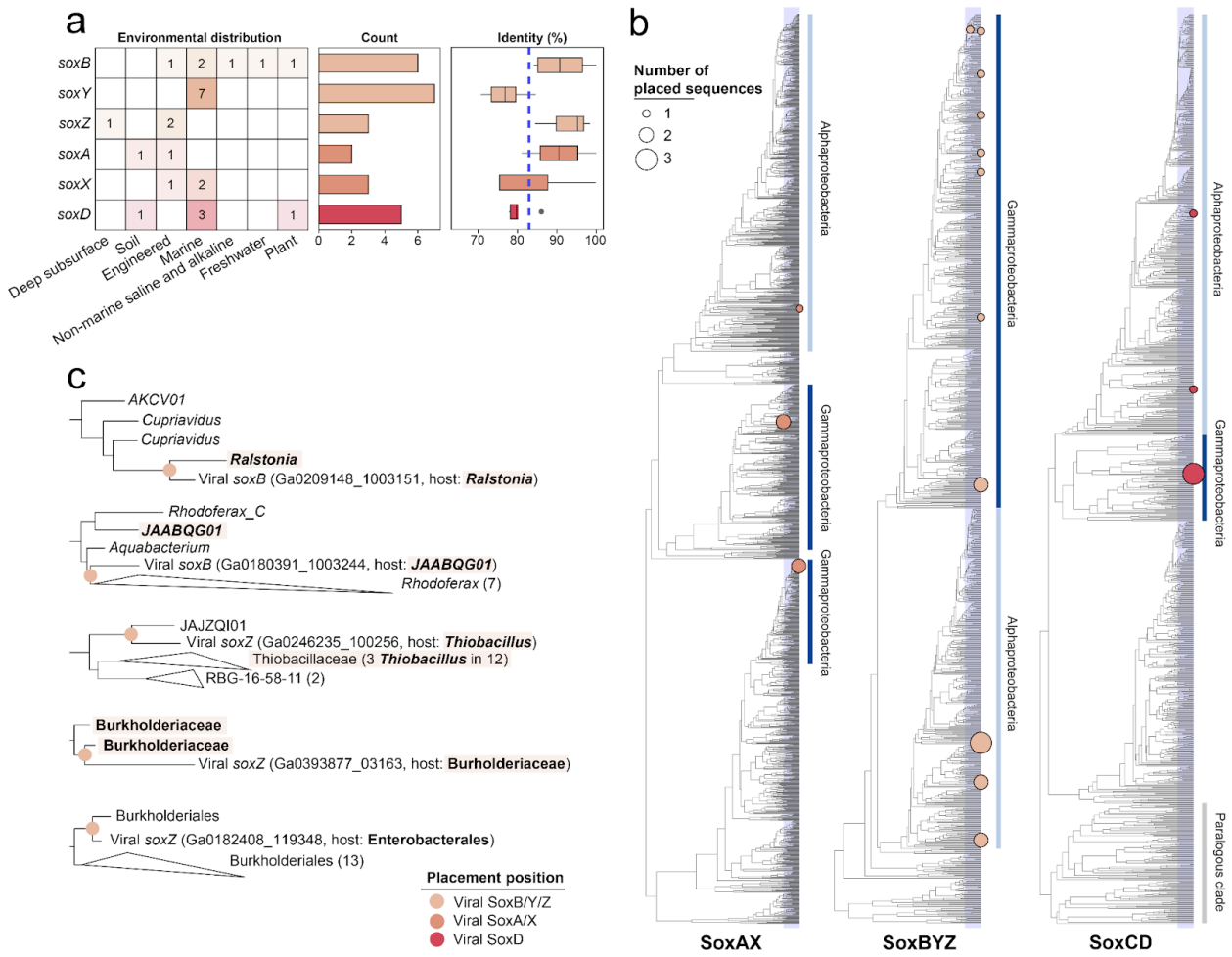

Figure 5 | Dissemination of sox genes across viruses. (a) Environmental distribution and prevalence of viral sox genes. The nucleotide sequence identity of viral genes to their closest bacterial homologs are shown as boxplot. The central line, lower, and upper limit of the box indicate the median, 25th, and 75th percentile, respectively; the whiskers extend to 1.5 times of 25-75th interquartile range from the box edges. The blue dash line denotes the median sequence identity of all virus-host gene pairs. **(b)** Phylogenetic placement of viral sequences on the reference trees of bacterial Sox proteins. Circles represent inferred positions of viral sequences, with the size scaled by the frequency of viral sequences placed to the same node. Tree branches were adjusted to the relative evolutionary divergence (RED) value of each node, which ranges from 0 (root) to 1 (tips), with values closer to 1 indicating nodes closer to the tree's terminal branches. Nodes located in shaded backgrounds denote those with RED ≥ 0.9. Major bacterial taxa for each clade are labeled on the right side of the trees. **(c)** The detailed phylogenetic relationships between viral Sox proteins and homologs from the predicted virus hosts. Virus-host protein pairs forming phylogenetic neighbors are shown as rectangular subtrees.

Reviewer #2 (Remarks on code availability):

I did not look at the code in great detail, but checked that it was available. I did not attempt to run it. This code seems to be specifically for identifying sox, rdsr, and shdr homologs in GTDB, which was one part of the bioinformatics analysis conducted fort his manuscript.

Response: All code relevant to the analyses presented in the main text has now been uploaded to the GitHub repository (https://github.com/RHarperR/SoxEvolution), including the newly added

scripts used for the additional analyses described in the SI. The repository has been organized to ensure that all main results in our manuscript can be readily traced and reproduced.

References:

Boughanemi, S. *et al*. Microbial oxidative sulfur metabolism: biochemical evidence of the membrane-bound heterodisulfide reductase-like complex of the bacterium *Aquifex aeolicus*. *FEMS Microbiol. Lett.* **363**, fnw156 (2016).

Davín, A. A. *et al*. A geological timescale for bacterial evolution and oxygen adaptation. *Science* **388**, eadp1853 (2025).

Hensen, D., Sperling, D., Trüper, H. G., Brune, D. C. & Dahl, C. Thiosulphate oxidation in the phototrophic sulphur bacterium *Allochromatium vinosum*. *Mol. Microbiol*. **62**, 794–810 (2006).

Holkenbrink, C., Barbas, S. O., Mellerup, A., Otaki, H. & Frigaard, N.-U. Sulfur globule oxidation in green sulfur bacteria is dependent on the dissimilatory sulfite reductase system. *Microbiology* **157**, 1229–1239 (2011).

Martinez-Gutierrez, C. A. & Aylward, F. O. Phylogenetic signal, congruence, and uncertainty across bacteria and archaea. *Mol. Biol. Evol.* **38**, 5514–5527 (2021).

Martinez-Gutierrez, C. A., Uyeda, J. C. & Aylward, F. O. A timeline of bacterial and archaeal diversification in the ocean. *eLife* **12**, (2023).

Nishihara, A., Tsukatani, Y., Azai, C. & Nobu, M. K. Illuminating the coevolution of photosynthesis and bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **121**, e2322120121 (2024).

Response to Reviewer's Comments

Reviewer #2 (Remarks to the Author):

The authors have satisfactorily addressed all of my comments and concerns. The detailed explanation of the calibration points is particularly helpful, as added to the supplementary material. They appear to have added extra calibration points and new models of evolution and conducted extensive sensitivity testing to demonstrate that their results are robust to the evolutionary model and underlying assumptions, and they have conducted formalized gene tree-species tree reconciliation that shows their results are consistent with the parsimony analysis. I commend the authors for their nice work on this paper.

Response: We are deeply grateful to the reviewer for their positive feedback and highly encouraging comments. We sincerely appreciate the reviewer's constructive guidance and insightful suggestions throughout the review process, which have significantly improved the quality and robustness of our manuscript.

Thank you again for your time, effort, and recognition of our work.