RESEARCH ARTICLE SUMMARY

EVOLUTION

A geological timescale for bacterial evolution and oxygen adaptation

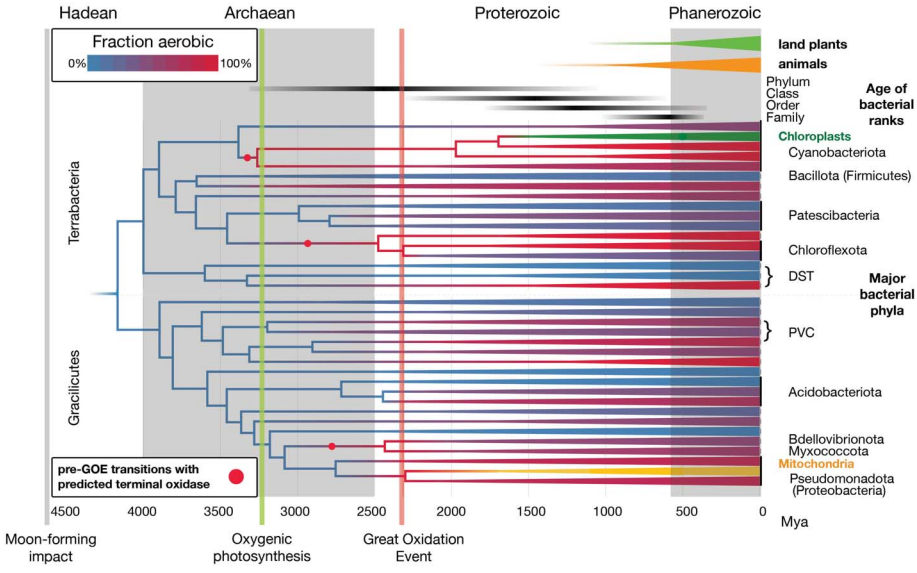
Adrián A. Davín*, Ben J. Woodcroft*, Rochelle M. Soo, Benoit Morel, Ranjani Murali, Dominik Schrempf , James W. Clark, Sandra Álvarez-Carretero, Bastien Boussau, Edmund R. R. Moody, Lénárd L. Szánthó, Etienne Richy, Davide Pisani, James Hemp, Woodward W. Fischer, Philip C. J. Donoghue, Anja Spang, Philip Hugenholtz*, Tom A. Williams*†, Gergely J. Szöllősi*†

INTRODUCTION: Microbial life dominates the biosphere, but a timescale of early microbial evolution has proven elusive as a result of an inadequate fossil record. The lack of maximum age calibrations—the earliest point in time at which a given group might have emerged—is particularly problematic. However, the geochemical record bears the imprint of microbial metabolism through time, providing a complementary source of information. A pivotal event in this history was the Great Oxidation Event (GOE) ~2.43 to 2.33 billion years ago (Ga), which marked a substantial increase in atmospheric oxygen. This transition, driven by the evolution of cyanobacterial oxygenic photosynthesis and carbon burial, transformed the biosphere from predominantly anoxic to oxic, causing widespread adaptation to oxygen. In this study, we used the temporal link between atmospheric oxygenation and the evolutionary spread of aerobic metabolism to calibrate the phylogeny of the bacterial domain.

RATIONALE: To date the bacterial tree, we introduced multiple new maximum age calibrations by linking the GOE to the age of aerobic lineages. We used a Bayesian approach that assumes that aerobic nodes are unlikely to be older than the GOE but can predate it given sufficient evidence from fossils or sequence divergence. To implement this approach, we integrated phylogenetic reconciliation with machine learning to map transitions from anaerobic to aerobic lifestyles onto the bacterial tree. By aggregating signals across the genome, we could robustly infer aerobic and anaerobic phenotypes from incomplete ancestral gene repertoires.

RESULTS: We identified 84 anaerobic to aerobic transitions on a species tree of 1007 bacteria. Most transitions occurred after the GOE and were driven by horizontal acquisition of respiratory and oxygen tolerance genes. However, despite the GOE calibration, at least three transitions predated this event, suggesting that aerobic respiration evolved before widespread atmospheric oxygenation and may have facilitated the evolution of oxygenic photosynthesis in cyanobacteria. Our molecular clock analyses estimated that the last bacterial common ancestor lived in the Hadean or earliest Archaean era (4.4 to 3.9 Ga), whereas bacterial phyla originated in the Archaean and Proterozoic eras (2.5 to 1.8 Ga); most bacterial families are as old as land plants and animal phyla, dating back to the late Proterozoic (0.6 to 0.75 Ga).

CONCLUSION: We infer that the earliest aerobic bacteria emerged in the Archaean, predating the GOE by 900 million years. After the GOE, aerobic lineages experienced faster diversification than their anaerobic counterparts, highlighting the impact of atmospheric oxygenation on bacterial evolution. The approach developed here provides a framework for linking microbial traits to Earth's geochemical history, offering a pathway for exploring the evolution of other phenotypes in the context of Earth's history. ■



An integrated approach to date bacterial evolution and reconstruct the history of oxygen adaptation. We inferred a bacterial timetree by integrating genomic, fossil, and geochemical data and linking oxygen tolerance and aerobic metabolism to the GOE. Colors denote anaerobic (blue) and aerobic (red) states, whereas shades of purple show the fraction of aerobic lineages within extant bacterial phyla. Mitochondria and plastids were included to leverage the more extensive eukaryotic fossils. Land plants and animals are indicated for temporal comparison.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: aaredav@gmail.com (A.A.D.); b.woodcroft@qut.edu.au (B.J.W.); p.hugenholtz@uq.edu.au (P.H.); thwillia@gmail.com (T.A.W.); gergely.szollosi@oist.jp (G.J.S.)
†These authors contributed equally to this work.
Cite this article as A. A. Davín *et al.*, *Science* 388, eadp1853 (2025). DOI: 10.1126/science.adp1853

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.adp1853

RESEARCH ARTICLE

EVOLUTION

A geological timescale for bacterial evolution and oxygen adaptation

Adrián A. Davín[1,2,3]*, Ben J. Woodcroft[4]*, Rochelle M. Soo[1], Benoit Morel[5,6], Ranjani Murali[7], Dominik Schrempf[2,8], James W. Clark[9,10], Sandra Álvarez-Carretero[9], Bastien Boussau[11], Edmund R. R. Moody[9,12], Lénárd L. Szánthó[2,13,14], Etienne Richy[12], Davide Pisani[9,12], James Hemp[15], Woodward W. Fischer[7], Philip C. J. Donoghue[9], Anja Spang[16,17], Philip Hugenholtz[1]*, Tom A. Williams[12]*†, Gergely J. Szöllősi[2,8,13,14]*†

Microbial life has dominated Earth's history but left a sparse fossil record, greatly hindering our understanding of evolution in deep time. However, bacterial metabolism has left signatures in the geochemical record, most conspicuously the Great Oxidation Event (GOE). We combine machine learning and phylogenetic reconciliation to infer ancestral bacterial transitions to aerobic lifestyles, linking them to the GOE to calibrate the bacterial time tree. Extant bacterial phyla trace their diversity to the Archaean and Proterozoic, and bacterial families prior to the Phanerozoic. We infer that most bacterial phyla were ancestrally anaerobic and adopted aerobic lifestyles after the GOE. However, in the cyanobacterial ancestor, aerobic metabolism likely predated the GOE, which may have facilitated the evolution of oxygenic photosynthesis.

Establishing evolutionary timescales requires relaxed molecular clocks (*1*, *2*) in which rates of molecular evolution (the "ticking" of the clock) can be calibrated for each lineage using fossils of known ages that are assigned to the internal nodes of an evolutionary tree (*3*, *4*). The difficulty in estimating a dated tree (time tree) for Bacteria is a result of the dearth of fossil evidence for constraining the ages of clades, with maximum age calibrations being a particular problem. The only credible maximum for the great majority of lineages is the Moon-forming impact (MFI) 4.52 billion years ago (Ga) (*5*, *6*), which would have effectively sterilized the planet (*7*, *8*).

Given the paucity of fossil calibrations, traces of biological activity preserved in the sedimentary geochemical record provide an alternative source of information on the age of key lineages and metabolisms (*9*, *10*). Arguably the most notable event recorded in the geochemical record is the Great Oxidation Event (GOE), when oxygen began to accumulate in Earth's atmosphere ~2.43 to 2.33 Ga (*11*–*13*). The GOE was ultimately made possible by the emergence of a new prokaryotic metabolism, oxygenic photosynthesis, which geochemical evidence shows had likely evolved by ~3.22 Ga and has been attributed to Cyanobacteria (*14*–*17*). The consensus is that, despite the evolution of oxygenic photosynthesis some 900 million years (Myr) prior, most of life's diversity was anaerobic before the GOE (*18*), although local concentrations of oxygen may have been higher in some environments (*12*, *19*), and the extent to which aerobic life might have evolved and persisted before the GOE is debated (*19*–*23*). As oxygen levels rose, anaerobic life either retreated to anoxic niches or adapted to its presence, giving rise to a wide variety of lifestyles that could either tolerate or directly use oxygen in the respiratory chain (*24*). Most transitions to aerobic lifestyles would have happened during or after the GOE, potentially providing a maximum age for most oxygen-adapted lineages that can help to resolve the bacterial time tree. We develop and validate a probabilistic approach to date bacterial evolution using the GOE as a soft maximum age calibration on oxygen-adapted lineages (Fig. 1).

Results

Inferring the evolution of oxygen adaptation on the bacterial tree

We inferred a species tree of Bacteria using 1007 bacterial genomes representing most orders in the Genome Taxonomy Database (GTDB) (*25*) [Fig. 2A and supplementary materials (SM) section 1]. We used machine learning to predict oxygen adaptation from genomic gene content, finding that the gradient boosting method XGBoost (*26*) was the most accurate and robust, based on annotations from the BacDive database (see fig. S1 and table S1). Aerobic organisms, as defined in BacDive, are those that can grow in the presence of oxygen. This category includes both organisms that utilize oxygen as part of their respiratory chain and those that can simply tolerate its presence.

To identify the branches on the tree where transitions between anaerobic and aerobic lifestyles occurred, we reconstructed ancestral gene contents for all internal nodes of the species phylogeny using the phylogenetic reconciliation method ALE (*27*, *28*) (see SM section 2) and applied the XGBoost classifier to these nodes to infer the ability of ancestral organisms to grow in the presence of oxygen (SM section 3). In our dataset, we inferred 84 transitions from anaerobic to aerobic lifestyles, with 36% (30 transitions) occurring in the terminal leaves. Conversely, 66 transitions to anaerobic were observed, with 60% (41 transitions) affecting the terminal leaves.

Since ancestral gene contents can only be inferred with substantial uncertainty, we explored the impact of false positives and negatives (spuriously present or absent gene families) on our ancestral state inferences. Even with a joint 50% false negative and false positive rate, XGBoost retained >90% accuracy (fig. S1). This robustness implies that many gene families distinguish aerobes from anaerobes. Genes with the largest effect on the XGBoost predictions are predicted to be directly involved in aerobic respiration (such as terminal oxidases and the biosynthesis of their cofactors) and hallmark genes of anaerobic metabolism (such as pyruvate formate lyase; Fig. 2B).

To validate our findings, we built an alternative classifier based on only the trimmed amino acid concatenate used to infer the species tree (SM section 3). The concatenate-based classifier had a leave-one-out prediction accuracy of 83.7% and was in good agreement with the gene content-based classifiers for ancestral nodes (overall 90% with XGBoost; see SM section 3). This suggests that oxygen adaptation has left two consistent but distinct imprints in genomes, on gene content, and in the amino acid sequences of universal genes.

[1]The University of Queensland, School of Chemistry and Molecular Biosciences, Australian Centre for Ecogenomics, Brisbane, Queensland, Australia. [2]Department of Biological Physics, Eötvös Loránd University, Budapest, Hungary. [3]Department of Biological Sciences, Graduate School of Science, University of Tokyo, Tokyo, Japan. [4]Centre for Microbiome Research, School of Biomedical Sciences, Queensland University of Technology (QUT), Translational Research Institute, Woolloongabba, Australia. [5]Computational Molecular Evolution Group, Heidelberg Institute for Theoretical Studies, Heidelberg, Germany. [6]Institute for Theoretical Informatics, Karlsruhe Institute of Technology, Karlsruhe, Germany. [7]Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA, USA. [8]MTA-ELTE "Lendület" Evolutionary Genomics Research Group, Budapest, Hungary. [9]Bristol Palaeobiology Group, School of Earth Sciences, University of Bristol, Bristol, UK. [10]Milner Centre for Evolution, Department of Life Sciences, University of Bath, Bath, UK. [11]Laboratoire de Biométrie et Biologie Evolutive, Univ Lyon, Univ Lyon 1, CNRS, VetAgro Sup, Villeurbanne, France. [12]School of Biological Sciences, University of Bristol, Bristol, UK. [13]Institute of Evolution, Centre for Ecological Research, Budapest, Hungary. [14]Model-Based Evolutionary Genomics Unit, Okinawa Institute of Science and Technology Graduate University, Okinawa, Japan. [15]Metrodora Institute, West Valley City, UT, USA. [16]Department of Marine Microbiology and Biogeochemistry, NIOZ, Royal Netherlands Institute for Sea Research, Den Burg, Netherlands. [17]Department of Evolutionary & Population Biology, Institute for Biodiversity and Ecosystem Dynamics (IBED), University of Amsterdam, Amsterdam, Netherlands.
***Corresponding author. Email: aaredav@gmail.com (A.A.D.); b.woodcroft@qut.edu.au (B.J.W.); p.hugenholtz@uq.edu.au (P.H.); thwillia@gmail.com (T.A.W.); gergely.szollosi@oist.jp (G.J.S.)**
†These authors contributed equally to this work.

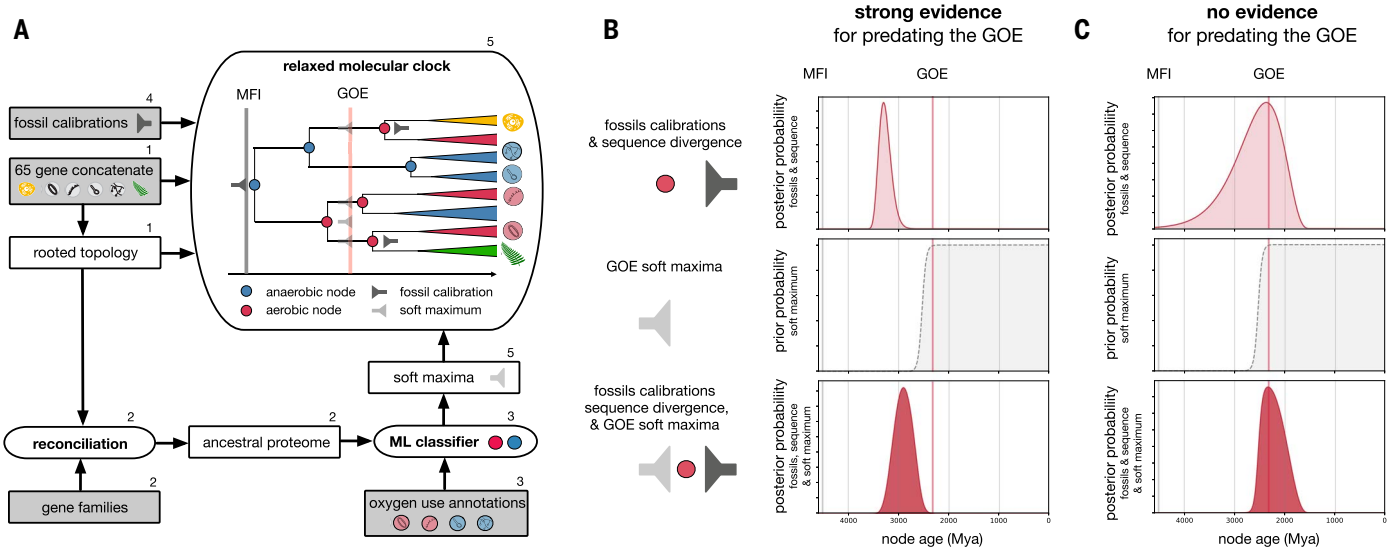


Fig. 1. An integrated approach to date bacterial evolution and reconstruct the history of oxygen adaptation. (**A**) We used relaxed molecular clock analyses to infer a dated phylogeny of bacteria, informed by several sources of information (gray boxes). These include orthologous marker genes (65-gene concatenate, including genes from the mitochondrial and plastid genomes, in yellow and green, respectively), calibrations, and ancestral transitions to oxygen use informed by phylogenetic reconciliation and machine learning (see supplementary material sections as indicated by numbers above boxes and main text). (**B** and **C**) Fossils provide minimum ages for clades (that is, they establish that a given clade existed at a certain point in the past), but establishing maximum age constraints is more challenging. In molecular clock analyses, the time information provided by fossils is encoded in the prior (*4*). Combined with the sequence information, the analysis results in a posterior estimate for the age of each node. In some cases (B), the available information from sequence data and fossils provides strong evidence that a node is truly ancient, but in the absence of strong evidence (C), node ages are skewed towards the root of the tree due to a lack of information, where the age of Earth (based on the MFI) provides the only maximum. (B) and (C) illustrate the effect of applying a soft maximum calibration at the GOE to aerobic nodes that represent these alternatives. In (B), the inferred node age remains old as sequence and fossil information overrule the soft maximum prior. In (C), sequence and fossil information do not provide compelling support for a pre-GOE age. Our analyses show that GOE soft maxima improve the accuracy of the inferred time tree with respect to an independent measure: the history of gene transfers among bacterial lineages (see later in main text and Table 1).

A bacterial time tree calibrated with the GOE

By identifying branches in our reference tree with different XGBoost-predicted aerobic states for ancestral and descendant nodes, we established a map of aerobic transitions on the bacterial phylogeny, to which the GOE soft maxima could be applied in dating analyses (Fig. 3). To infer the time scale of bacterial evolution, we expanded our species tree to include genes from eukaryotic organelles. Mitochondria and chloroplasts branch with Alphaproteobacteria (*29*, *30*) and Cyanobacteria (*31*, *32*), respectively, but their ages can be constrained using the eukaryotic fossil record, increasing the number of fossil calibrations available for dating the bacterial species tree (of 16 calibrated nodes, 10 are eukaryotic, with the eukaryotic record providing 5 of 7 nonredundant maximum age calibrations (SM section 4) (*33*, *34*). Since plants and algae have both mitochondria and chloroplasts, the same eukaryotic species divergences occur twice in the bacterial species tree, for example, the divergence between red and green algae appears in both the mitochondrial and plastid clades. These equivalent nodes can be "braced" (fixed to the same unknown age) (*33*, *34*) in the dating analysis, providing valuable additional calibration information (Fig. 3).

We then performed Bayesian relaxed clock-dating analyses to estimate evolutionary rates and divergence times and used a Gamma distribution to model the manifest rate variation among lineages in our focal analyses (shown in Fig. 3). We performed analyses under three different calibration strategies: "MFI," in which the MFI is used as the only calibration, providing a maximum age for bacterial evolution; "fossils," which includes the MFI calibration and fossil and geochemical calibrations on 15 additional nodes (see fig. S2 and SM section 4), and "fossils + GOE" as for fossils with an additional GOE soft maximum on the age of aerobic nodes. Note that, although transitions to oxygen were likely more common after the GOE, it is plausible that at least some aerobic lineages evolved beforehand (*19*–*22*, *35*–*37*). To account for this possibility, we implemented a soft maximum in the Bayesian analysis. Soft maxima have traditionally been used to reflect uncertainty in interpreting absence data in the fossil record (*38*), but they also provide a natural framework for the hypothesis that most aerobic transitions occurred after the GOE. Reflecting this expectation, we reduce the probability that aerobic nodes are older than the GOE, but they can nonetheless predate the GOE if there is sufficient evidence from fossils or sequence divergence to support it (Fig. 1, B and C, and figs. S3 and S23).

The impact of the GOE soft maximum is generally a shift in the origin of aerobic bacterial lineages toward the present, with the mean age of bacterial phyla becoming 328 Myr younger when this calibration is applied. It is important to note that this result is not predetermined by our assumption of a soft maximum at the GOE because, as noted above, the data (sequences and calibrations) can overcome the prior (Fig. 1B). This was indeed the case for five aerobic nodes (Figs. 3A and 5A), including the immediate antecedents of photosynthetic Cyanobacteriia (Fig. 3 transition I, Fig. 5B node 2, and SM section 5), consistent with geological evidence of at least some free oxygen at 3.225 Ga (see SM section 4).

Ancient gene transfers support the GOE soft maxima

We used the GOE soft maxima based on the prior hypothesis that aerobic lifestyles would have been more common after the GOE than before. However, given that some aerobic bacterial lineages predate the GOE in both the "fossils" and "fossils+GOE" analyses, it seems reasonable to ask whether applying the GOE soft maxima improves inference of the bacterial

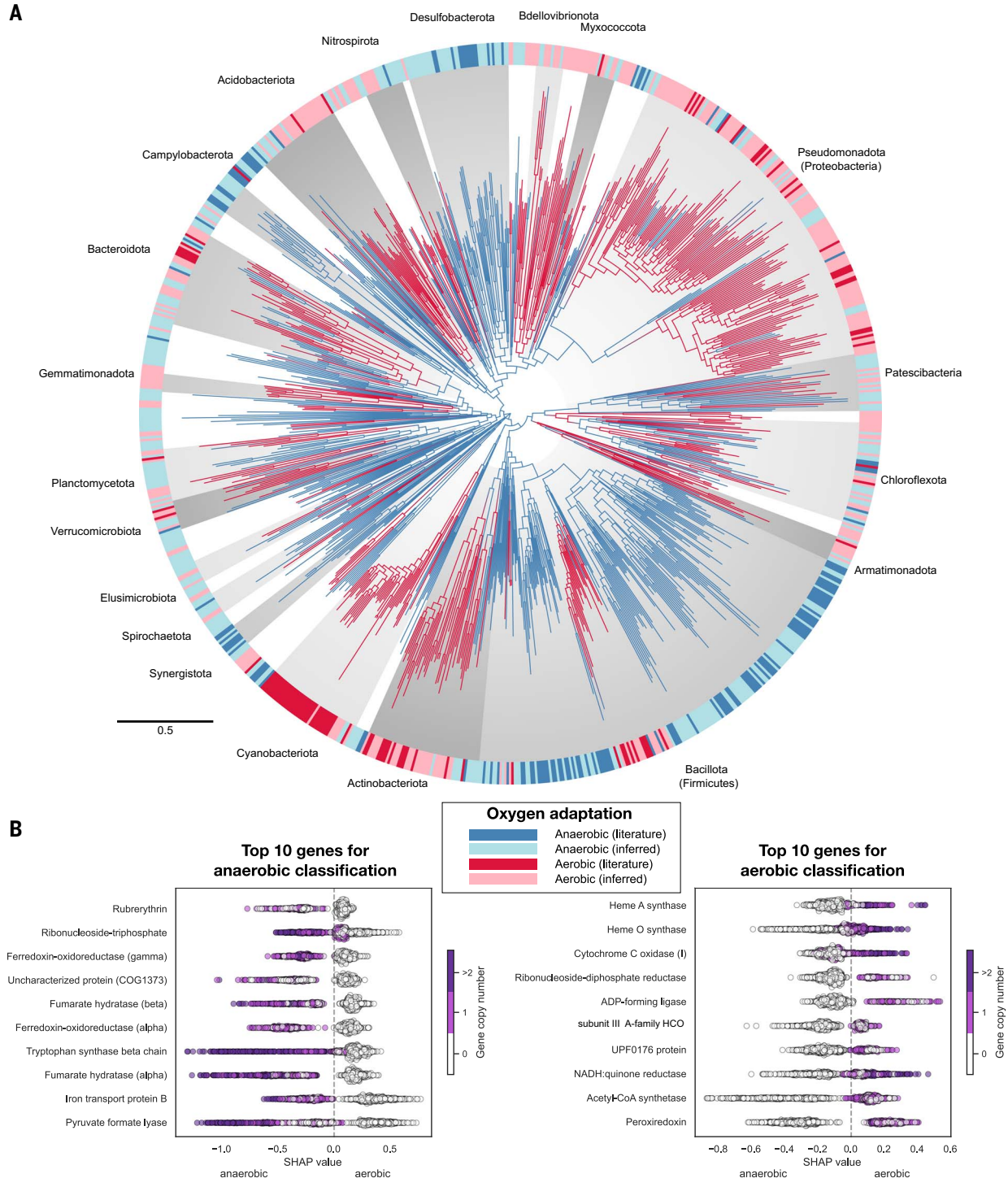


Fig. 2. The evolution of oxygen adaptation in Bacteria. (**A**) A species tree of Bacteria (inferred using 65 marker genes from 1007 genomes using a custom site-heterogeneous substitution model), with branches colored according to predicted ability to grow in the presence of oxygen (red, aerobic; blue, anaerobic). The outer ring shows observed (dark colors) or predicted (light colors) oxygen adaptation phenotypes for extant genomes. Major bacterial phyla (with >15 genomes in our tree) are named and delimited by shading. Branch lengths are proportional to the expected number of substitutions per site, as indicated by the scale bar. (**B**) Genes with the strongest contribution (SHAP values) to classification as aerobic or anaerobic according to the XGBoost classifier (see main text and SM). Each dot represents one node in the tree. On the left, genes with a higher copy number predict an anaerobic lifestyle; on the right, genes with a higher copy number predict an aerobic lifestyle. The genes most diagnostic of aerobic lifestyles include core components of aerobic metabolism such as HCO, enzymes for the biosynthesis of heme A, the cofactor of the most widely distributed O_2 reductase (A-family) from the HCO superfamily.

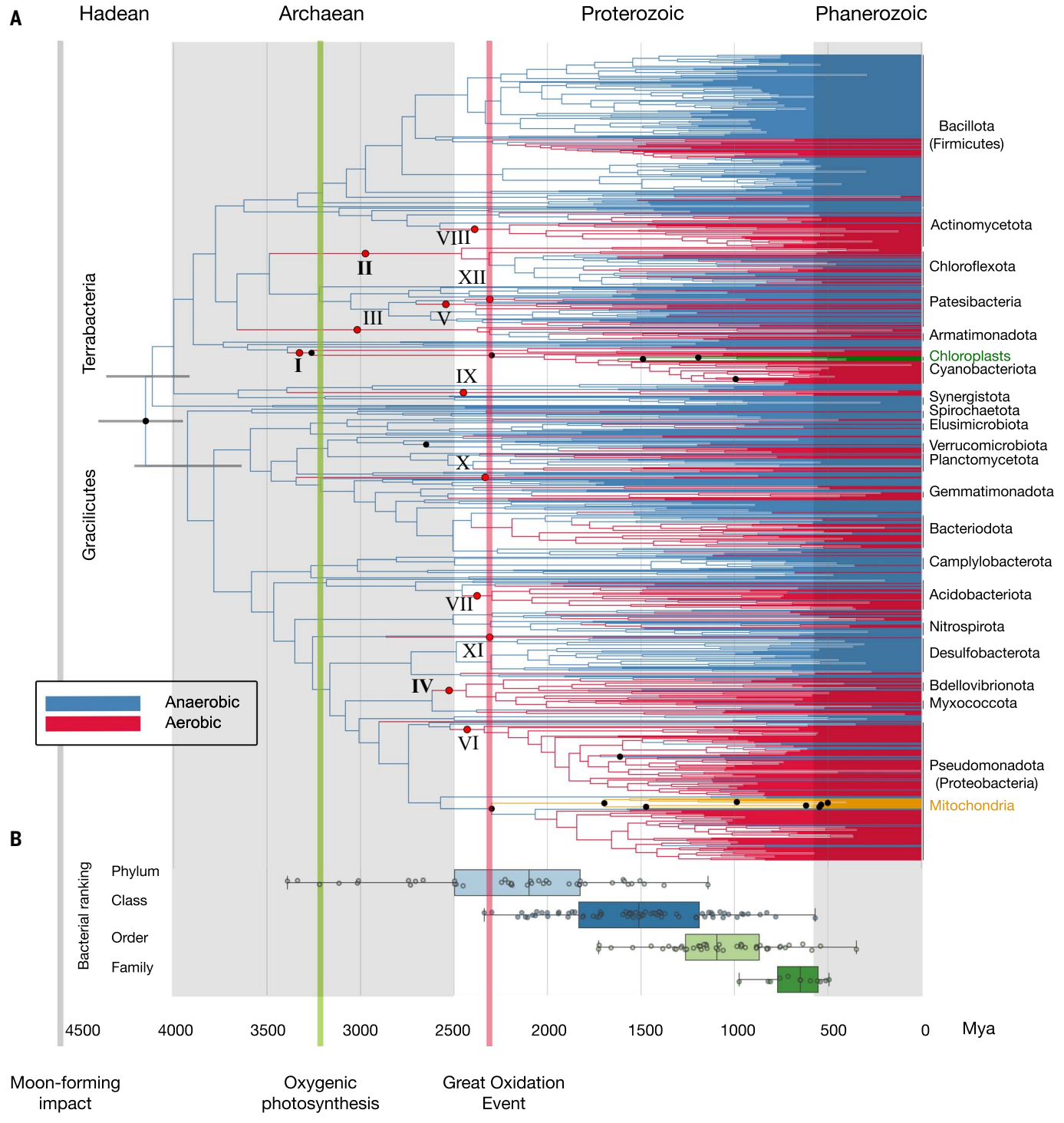


Fig. 3. A dated phylogeny of Bacteria. (**A**) Dated phylogenetic tree of Bacteria. Branch colors represent anaerobic (blue) and aerobic lineages (red). Mitochondrial and chloroplast branches are colored in orange and green, respectively. Branch lengths are proportional to geological time (in billions of years.) Horizontal gray bars indicate 95% highest posterior densities, shown for the root as well as the two major bacterial divisions, Gracilicutes and Terrabacteria. The vertical gray line indicates the MFI at ~4.52 Ga; the green line at ~3.23 Ga reflects the presence of fossil and isotopic evidence for oxygenic photosynthesis; the red line at ~2.33 Ga reflects the end of the GOE (see calibrations in SM section 4). Black dots denote nodes that were directly calibrated in molecular clock analyses. Red dots labeled with roman numerals indicate branches with inferred aerobic transitions that likely predate the GOE (Fig. 2). These include the MFI at 4.5 Ga as a maximum on Bacteria, three cyanobacterial fossils, the pigment okenane (diagnostic of Chromatiaceae), and 10 calibrations from the eukaryotic fossil record mapped to mitochondria and chloroplasts (see fig. S20). Alternating background shading distinguishes geological eons. (**B**) Inferred ages of taxonomic ranks as defined in GTDB. The names of phyla represented by eight or more genomes are shown. We show the age of the last common ancestors of major taxa (phyla, class, order, and family) represented in our 1007 genomes dataset by at least three genomes. The median ages are: phyla, 2115 Ma; class, 1524 Ma; order, 1120 Ma; family, 655 Ma. A chronogram with annotations is included in the supplementary data.

time tree. We reasoned that the histories of gene families carry phylogenetic signals to distinguish the dated trees because gene transfers can only occur between species that exist at the same time (*39*–*42*). To harness this information, we used a reconciliation model that only allows transfer between branches that overlap in time (see SM section 6). The likelihood of gene families under the reconciliation model provide a metric for choosing among different time trees (fig. S5), much as sequence likelihood can be used to choose among tree topologies in traditional phylogenetics (*43*–*45*).

To assess the effect of the GOE soft maximum on the time tree, we compared the support for the inferred timescales using 4713 gene families from the 1007 genomes (Table 1). As expected, the time tree calibrated with only the MFI had the lowest likelihood—that is, was least compatible with gene family histories. Calibrating the tree with fossils substantially improved the likelihood, whereas combining fossils with the GOE soft maximum further increased the likelihood and produced the best results. The improvement in likelihood was greatest when using the XGBoost predictor, though a substantial improvement in likelihood was observed for all predictors of oxygen adaptation tested (Table 1 and fig. S5), including the second-best ExtraTrees and the concatenate-based method (SM section 3).

A revised chronology of bacterial evolution

Our analyses support the view that the deepest divergences within the bacterial domain occurred early in Earth's history (Fig. 3). We estimate that the last bacterial common ancestor (LBCA; Fig. 4A) existed 4.4 to 3.9 Ga, which is after the MFI but likely before the period associated with the Late Heavy Bombardment (4.0 to 3.8 Ga) (*46*). The inferred age of LBCA is broadly consistent with recent analyses incorporating a wide sample of known bacterial and archaeal diversity (*22*, *47*, *48*), but older than some earlier analyses (*8*, *49*). These inferences suggest that the early Earth was more hospitable for prokaryotic life than has sometimes been thought (*7*). A frequent difficulty in deep-time molecular clock analyses is that age estimates are sensitive to the root maximum age calibration, with the inferred root ages clashing against the maximum root age due to a lack of informative maxima elsewhere in the tree (*34*, *50*–*52*). Our analyses incorporating different datasets and alternative calibrations show that the GOE soft maximum calibration provides new timing information that makes inferences robust to the model choice and root maximum age (Fig. 4B and SM section 5). For example, when setting the root maximum to 3.8 Ga (corresponding to the estimated end of the Late Heavy Bombardment), the age of LBCA is constrained against the maximum, but stabilizes at 4.4 to 3.9 Ga with older root maxima. Thus, the sequence data and fossils do not support the view that extant bacterial life originated after the Late Heavy Bombardment, even with the introduction of a GOE soft maxima on aerobic nodes.

We infer that the major descendant clades Gracilicutes and Terrabacteria radiated 4.2 to 3.6 Ga and 4.3 to 3.9 Ga, respectively, much earlier than the original estimates (*49*), but in broad agreement with some recent inferences (*22*). The extant phyla with the oldest crown groups are Bacillota (formerly Firmicutes; 3.6 to 3.0 Ga), Patescibacteria (3.4 to 2.9 Ga), Actinomycetota (3.4 to 2.8 Ga) and Cyanobacteriota (3.5 to 3.3 Ga). Pseudomonadota (formerly Proteobacteria; crown age 2.9 to 2.6 Ga) are somewhat younger than Cyanobacteriota (3.5 to 3.3 Ga) and other major terrabacterial phyla despite their extant phylogenetic and metabolic diversity. We inferred the age of extant oxygenic photosynthesizers, crown Cyanobacteriia, to be 2.5 to 2.1 Ga. This date has varied in previous analyses (*22*, *41*, *53*), ranging from the Proterozoic back to the Archaean (3.9 to 1.7 Ga). In our dating analysis we applied the GOE soft maximum to all aerobic nodes, including crown oxygenic photosynthetic Cyanobacteriia. However, since this ancestor was already able to make its own oxygen through photosynthesis, we also performed an analysis excluding crown Cyanobacteriia from the GOE soft maximum calibration, resulting in a moderately older estimate (2.7 to 2.1 Ga, fig. S28), which is in broad agreement with recent analyses using alternative dating approaches (*53*).

We find that the relative evolutionary divergence (RED) metric—which measures genetic divergence between sequenced taxa and is used by GTDB to partition prokaryotic diversity into taxonomic groups of similar age—correlates well with absolute geological time, with a Spearman correlation coefficient (ρ) of 0.93 between RED values and the age of nodes (SM section 5). A previous study focusing on fungal taxonomy found a similar consistency between RED and relative divergence times (*54*), suggesting that RED is generally a good proxy for divergence time, despite its relative simplicity.

Our analysis makes plain the enormous difference in evolutionary timescales for prokaryotic and eukaryotic diversity and the temporal heterogeneity of Linnean ranks (*55*, *56*). For example, we estimate that the last common ancestor of extant mitochondria (and therefore of extant eukaryotes) lived 1880 to 1502 million years ago (Ma). The eukaryotic domain of life is therefore comparable in age to a bacterial class whereas previous estimates of the crown ages of groups such as metazoans (animals) 833 to 650 Ma (*57*, *58*) and embryophytes (land plants) 515 to 494 Ma (*59*) correspond roughly to bacterial families whose ages range from 899 to 457 Ma in our analysis (Fig. 3). Indeed, many modern bacterial families were already established in the Archaean or Proterozoic (see Fig. 3). Furthermore, Fig. 3 reveals that a decisive fraction of extant bacterial diversity evolved before the emergence of bacterial phyla, which supports the recent proposal to add the rank of kingdom between domain and phylum (*60*).

Table 1. Comparison of likelihoods under the time-constrained DTL model. Using different consensus chronograms, 4713 families were reconciled. We show both the total log-likelihood (resulting from adding all the log-likelihoods from every reconciliation) and the difference with the least likely chronogram (MFI is the only calibration implemented as a maximum age calibration on the root). The analysis indicated that the dated tree incorporating the XGBoost predictions of oxygen use was significantly better (AU test $P < 0.05$) (*93*) than all others. Chronograms are included in the supplementary data.

Condition	Log-likelihood	Improvement over MFI	AU test *P*-value
MFI	−4736205	0	1×10^{-14}
MFI + Fossils	−4735968	237	8×10^{-17}
MFI + Fossils + GOE (Concatenate)	−4735871	334	2×10^{-8}
MFI + Fossils + GOE (ExtraTrees)	−4735539	666	9×10^{-6}
MFI + Fossils + GOE (XGBoost)	−4735472	733	1

Aerobic metabolism predated the GOE

Returning to the history of aerobic adaptation across the bacterial tree, LBCA and the common ancestors of Gracilicutes and Terrabacteria are inferred to be anaerobic. Indeed, most (38 out of 49) bacterial phyla are inferred to be ancestrally anaerobic. For example, our analysis identifies two early independent aerobic transitions within the Pseudomonadota (Proteobacteria), consistent with published gene trees showing multiple horizontal acquisitions of key respiratory enzymes in this clade (*61*).

We find that most aerobic transitions happened after the GOE, in both the analysis with (72 of 84 post-GOE, Fig. 3A) and without the

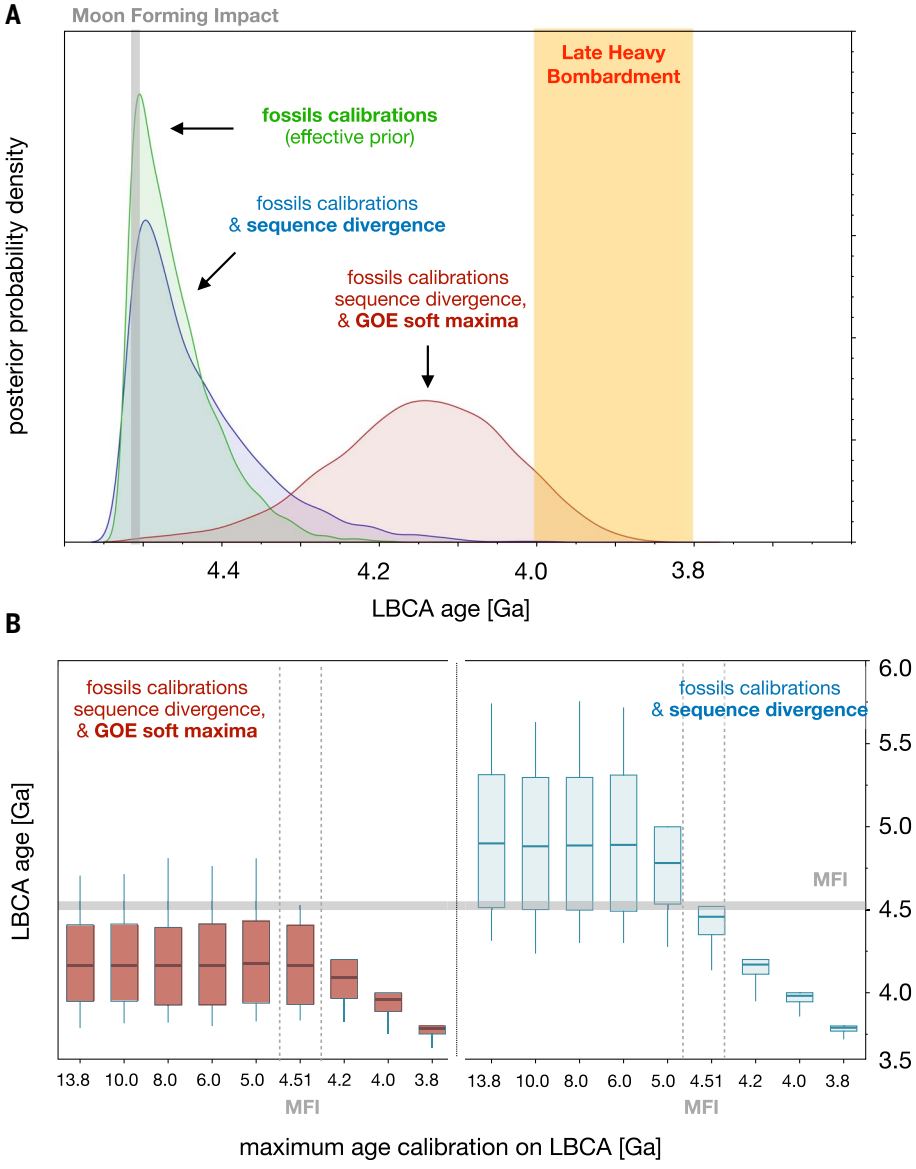


Fig. 4. The inferred age of the LBCA under a range of analysis conditions. (**A**) Inferred age of LBCA relative to the MFI and the period associated with the Late Heavy Bombardment. An analysis considering only the tree topology and calibrations, including the LBCA maximum root age based on the MFI (that is, the effective prior, in green) places most of the probability close to the root maximum age calibration at 4.52 Ga. The effect of adding sequence data is to shift the age range for LBCA to be moderately younger, albeit still abutting the MFI (blue). Adding the additional soft maxima on aerobic nodes (the GOE soft maxima, red) ameliorates the very old age suggested by the tree topology and calibrations by themselves. (**B**) Effect of different root age maxima on the inferred age of the last bacterial common ancestor, with and without the GOE soft maxima. When the GOE soft maximum calibration is applied (red, on left) the inferred age of LBCA becomes robust to choices of root age maxima that extend back to the age of the known universe at 13.8 Ga, likely due to the additional rate information provided by the new maxima. Considering only the calibrations, root maximum, and sequence data (blue to the right), the inferred age of LBCA is sensitive to the choice of root maximum. For example, when maxima are older than the MFI, the age of LBCA exceeds the age of Earth. This effect has been discussed before (*8*, *47*, *50*) and may reflect a lack of maximum ages to calibrate the molecular clock early in life's history.

GOE soft maximum (61 of 84 post-GOE, fig. S23). However, even with the GOE soft maximum, the median posterior age of 12 transitions predated the GOE, of which 5 occurred pre-GOE with 95% posterior probability or greater (Fig. 5A), providing evidence for oxygen adaptation before the sustained oxygenation of the atmosphere (*20*, *23*, *35*, *62*–*65*). Pre-GOE, microorganisms might have encountered localized pockets of oxygen produced abiotically by, for example, photolysis or radiolysis of water (*65*, *66*), crushed silicate rock-water reactions (*67*–*69*) or local concentrations of oxygen produced in microbial mats or marine oxygen oases following the evolution of oxygenic photosynthesis (*70*–*72*).

The earliest inferred aerobic transition occurred by 3.2 Ga, that is, at least 900 Myr before the GOE, in the common ancestor of the cyanobacterial classes Cyanobacteriia and Vampirovibrionia (henceforth CV ancestor, node 2 in Fig. 5B). To investigate whether this lineage was already capable of aerobic respiration, we developed a second gene content–based classifier that predicts the ancestral presence of terminal oxidases [i.e., 22 heme-copper oxygen reductases (HCOs) and cytbd families and subfamilies] based on reconstructed ancestral gene content (see SM section 3). We first applied the two classifiers for oxygen adaptation (aerobe/anaerobe) and terminal oxidases (presence/absence) to the 1007 genomes of extant organisms to sanity check the four possible combinations. As expected, almost all modern aerobes were predicted to have terminal oxidases (~97%, 433 of 445 genomes); however, the Patescibacteria were a conspicuous exception in that they include predicted aerobes that lack terminal oxidases (12 of 445 genomes). This may be due to their host-associated lifestyle in which they indirectly profit from the terminal oxidase of an aerobic host (*73*). For example, the archaeal DPANN symbionts belonging to the Nanohaloarchaeota include aerotolerant organisms that depend on aerobic hosts (*74*, *75*). Notably, almost half of contemporary anaerobes were predicted to have terminal oxidases (~46%, 261 of 562 genomes, see SM section 3 for manual verification of these cases), which may be used for oxygen detoxification (*76*–*78*). The remaining half were predicted to be anaerobes lacking terminal oxidases, which is also the inferred state of most deep ancestral nodes.

The terminal oxidase classifier predicted that three of the five high-confidence pre-GOE aerobic transitions involved terminal oxidases: those of early Cyanobacteriota (94% probability), Chloroflexota (89%), and Myxococcota (99%; transitions I, II, and IV in Fig. 5A). The combination of the signal for oxygen adaptation and for the ancestral presence of terminal oxidase provides compelling support for the hypothesis that the CV ancestor was already able to use oxygen. However, the probabilities for the presence of individual components of the respiratory chain are only moderate in the CV ancestor (Fig. 5B; see also fig. S8), and thus machine learning classifiers that aggregate data across the genome provide essential additional evidence. This finding contrasts with our previous inference that the CV ancestor was anaerobic based solely on the presence of different terminal oxidase types in extant

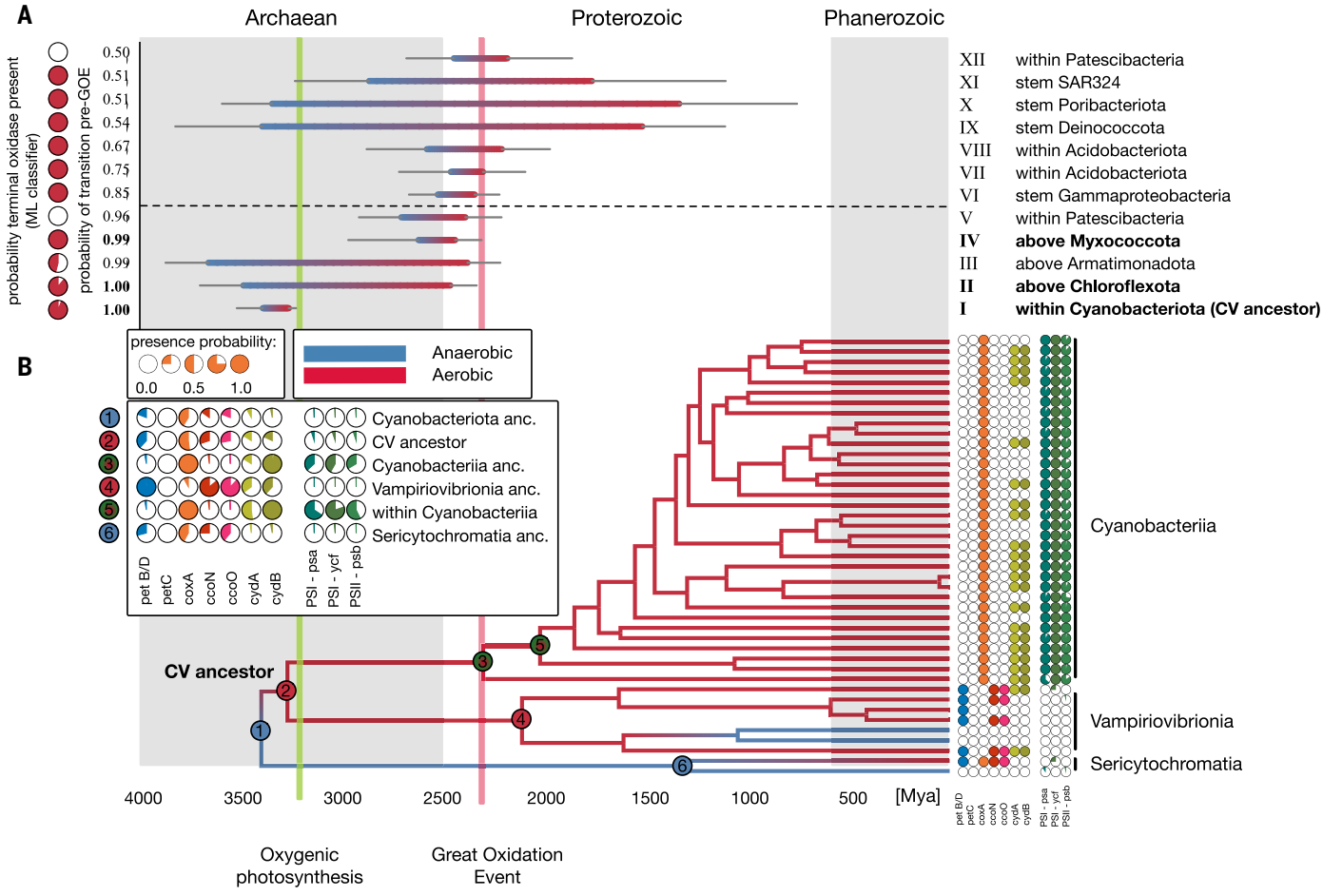


Fig. 5. Aerobic metabolism predates the GOE and likely enabled oxygenic photosynthesis. (**A**) Age ranges for inferred aerobic transitions (indicated by a shift from blue to red) that likely predate the GOE. The depicted age ranges extend from the oldest 5% tail of the anaerobic ancestor to the youngest 5% tail of the aerobic descendant. 12 transitions have a median transition age younger than the GOE (that is, pre-GOE transition probability >0.5); 3 of the 5 transitions that have >0.95 of their probability before the GOE were inferred to have a HCO/cytbd terminal oxidase using an ML classifier (names in bold, see also fig. S18). (**B**) Aerobic transitions and gene content evolution within Cyanobacteriota. We infer that the transition to an aerobic lifestyle occurred along the branch leading to the last common ancestor of Cyanobacteriia and Vampirovibrionia (node 2; see also fig. S4 and SM section 5), whereas oxygenic photosynthesis appears later, within the stem lineage leading to crown Cyanobacteriia (node 3; i.e., within total-group Cyanobacteriia), although its evolutionary origins are contested (*94*, *95*). Although the crown cyanobacterial ancestor is very likely to have been capable of oxygenic photosynthesis, presence probabilities for some of the key genes are only moderate as a result of the conservative nature of reconciliation analyses in the presence of phylogenetic noise (*80*). The XGBoost prediction is consistent with presence-absence patterns for Complex III and IV genes in extant and ancestral Cyanobacteriota, with some support for the hypothesis that CoxA was already present in the CV ancestor. The columns include the average probability of genes involved in photosynthesis, PS I - Psa contains photosystem I's primary reaction center; PS I - Ycf involves photosystem I assembly, and PS II - Psb forms the core of photosystem II, essential for oxygen evolution. Overall, these analyses are consistent with the hypothesis that the cyanobacterial lineage that evolved oxygenic photosynthesis was already capable of aerobic respiration (*21*, *62*). The green line at ~3.23 Ga reflects the presence of fossil and isotopic evidence for oxygenic photosynthesis; the red line at ~2.32 Ga reflects the end of the GOE (see calibrations in SM section 4). Alternating background shading denotes geological eons.

Cyanobacteriia and Vampirovibrionia (*17*, *61*), highlighting the additional insights provided by phylogenetic reconciliation combined with machine learning.

The comparison of results from the classifiers for oxygen adaptation and terminal oxidases suggests that terminal oxidases predate transitions to aerobic lifestyles, sometimes substantially, with some deep anaerobic nodes within Gracilicutes predicted to encode terminal oxidases with high confidence (fig. S18). This result is consistent with previous work suggesting that some key oxygen-using enzymes trace their origins prior to the GOE (*23*), to LBCA (*20*) or even the last universal common ancestor (*21*). One plausible scenario for the evolution of aerobic respiration is that the GOE and associated increase in atmospheric oxygen levels created a strong selective pressure for the evolution of new protein machinery that could harness this resource for cellular energy (*79*). However, our work [and that of some others (*20*, *21*, *23*, *35*, *63*)] instead suggests that aerobic respiration might have evolved more gradually, with the earliest terminal oxidases adapted to deal with trace amounts of oxygen produced initially by nonphotosynthetic means, as described above. If so, that would suggest that modern anaerobes that encode terminal oxidases but do not appear to use them as a means of energy conversion (*76*–*78*) might represent the best modern analogs of this early period of HCO evolution. Future research involving ancestral sequence reconstruction of the HCO in the CV ancestor might provide insight into the selective pressures that drove the evolution of this enzyme.

The first appearance of key oxygenic photosynthesis genes is in the Cyanobacteriia ancestor

(node 3 in Fig. 5), postdating aerobic metabolism in this lineage. This lends support to the hypothesis that oxygen adaptation was a necessary prerequisite for, and may have facilitated, the evolution of oxygenic photosynthesis, as it would have enabled early Cyanobacteriia to manage or even capitalize upon the oxygen produced (*21*, *62*–*64*). However, confidently determining the order of these events is challenging because demonstrating the absence of a trait becomes more difficult as phylogenetic signal is diluted further back in evolutionary time (*80*). Given recent work suggesting that phototrophy might have evolved earlier in bacterial evolution (*81*, *82*), we cannot reject the possibility that oxygenic photosynthesis predated, or coevolved with, aerobic metabolism in the cyanobacterial lineage. To explore the impact of these uncertainties on our time tree of bacteria, we performed clock analyses in which we did not assume that the organisms carrying out oxygenic photosynthesis at 3.2 Ga were stem-Cyanobacteriia (see SM section 4), instead using a conservative younger minimum age for the CV ancestor at the GOE (fig. S4). In this analysis, the CV ancestor was inferred to be younger (2.9 to 2.4 Ga), but our results were not otherwise majorly affected.

Tempo and mode of bacterial evolution

We estimated rates of gene content evolution in aerobes versus anaerobes by combining the dated tree with the gene tree-species tree reconciliations. Numbers of gene transfers and losses per million years that have left a phylogenetic signal, normalized by genome size, were broadly similar in aerobes and anaerobes, with about one transfer and one loss recorded per million years, while inferred rates of gene duplication were about 100 times lower. Gene transfer was instrumental in the spread of aerobic metabolism, including transfer of electron transport chain components and other genes associated with transitions to aerobic life (SM section 7).

To evaluate how bacterial gene repertoires have evolved over time, we applied nonmetric multidimensional scaling to quantify the disparity among the gene content of modern and ancestral bacteria, using the reconstructed ancestral gene repertoires as ancestral states (see SM section 7). Plotting disparity through time (Fig. 6A) shows that most variation in bacterial gene repertoires was already established prior to 3.5 Ga, during the radiation of the Terrabacteria and Gracilicutes clades, supporting an "Archaean expansion" of gene families (*83*).

Next, we investigated the impact of the transition to aerobic metabolisms on bacterial evolution. To compare the relative diversification rate of aerobes and anaerobes, we used a metric based on the time intervals between speciation events (Fig. 6B and SM section 7). This analysis suggests two main periods of oxygen adaptation: a rapidly established diversification advantage of aerobes beginning at the GOE (Fig. 6E) that was sustained for the following ~1.5 billion years, after which it waned but then rebounded during the last 500 Myr. A higher diversification rate for aerobes is also evident in a complementary metric that measures the directionality of transitions between aerobic and anaerobic metabolism (Fig. 6C and SM section 7). We interpret this pattern to reflect a combination of the creation of aerobic niches at the GOE and additional higher-oxygen aerobic niches, as levels of atmospheric free oxygen underwent a second sustained rise in the Paleozoic (*84*, *85*), and the selective extinction of early anaerobic lineages as a consequence of oxygenation, as well as reduction and fragmentation of their niche space. The gradual, rather than punctuated, rise in aerobic lineages throughout the Mesoproterozoic (Fig. 6D) is consistent with relatively low levels of oxygen during this period (*84*, *86*, *87*), which could have served to temper aerobic diversification.

In addition to the two periods during which aerobic lineages had a relative advantage, both diversification metrics also record a late reversion of the earlier trends, in which anaerobes are at an advantage ~200 Ma (Fig. 6). These inferences are uncertain because confidence intervals are wider in this period owing to the low number of speciation nodes. Nonetheless, it seems possible that this signal reflects adaptation to new anoxic niches, such as the gastrointestinal tracts of early animals or anaerobic niches within decaying biomass (*88*, *89*). This hypothesis could be tested in the future by expanding taxon datasets and analyzing bacterial lineages that have diversified into these anoxic niches using the methods developed in this study.

Discussion

Our analyses show a new way to link bacterial gene content to phenotype, and to trace those relationships back in time. By combining phylogenetic reconciliation with machine learning, we predicted the oxygen use phenotypes of ancestral nodes in the bacterial tree. We compensated for loss of phylogenetic resolution in deep time by aggregating the signal from many gene families both directly and indirectly associated with the phenotype. While the approach assumes that the genes predictive of aerobic metabolism in modern organisms are also predictive for their ancestors, our analyses suggest that the signal we capture is robust to substantial levels of noise, and so reliably tracks anaerobe-to-aerobe transitions in deep time. This allowed us to calibrate bacterial evolution to the record of biospheric oxygenation, greatly augmenting the limited fossil record of early life and bringing a new level of resolution to the study of evolution in deep time. Our time tree highlights the disparity between bacterial and eukaryotic classification: most extant bacterial phyla radiated in the Archaean and early Proterozoic, predating the eukaryotic domain of life. The earliest aerobes that left extant descendants lived in the Archaean (3.22 to 3.25 Ga), predating the GOE by about 900 Myr, and gave rise to the Cyanobacteriia and Vampirovibrionia, suggesting that aerobic metabolism evolved prior to oxygenic photosynthesis and the GOE in a restricted number of lineages. The GOE then drove an intense period of diversification and spread of aerobic metabolism, followed by a second phase of diversification at the end of the Proterozoic, consistent with a Mesoproterozoic redox stasis and an expansion of oxygenated habitats thereafter. This was enabled by the widespread transfer of respiratory complexes based on oxygen, and lineages that became aerobic subsequently experienced a higher diversification rate than contemporaneous anaerobes. Similar techniques could be applied to study the evolution of other phenotypic traits in light of Earth's history as well as to further investigate oxygen adaptation within the Archaea or both domains simultaneously to develop a broader picture of the biological response to atmospheric oxygenation on the early Earth. We anticipate that high-resolution time trees of microbial life will greatly enrich our understanding of the interplay between biological and geological evolution.

Materials and methods summary

Species tree inference

We inferred a rooted bacterial phylogeny using a 65-marker gene concatenate derived from 1007 genomes representatively sampled from the GTDB taxonomy (05-RS95). The tree was inferred using a new, dataset-specific, site-heterogeneous substitution model (LG+EDM64+G). We then added mitochondrial and plastid clades to the inferred species tree to make use of the eukaryotic fossil record in subsequent molecular clock analyses. For more details, see SM section 1.

Ancestral genome reconstruction

We reconstructed ancestral gene content by reconciling 4856 gene families against the species tree using the undated algorithm in amalgamated likelihood estimation (ALE) (*27*). To reduce phylogenetic noise, initial gene families based on Cluster of Orthologous (COG) annotations were divided into subclusters if long internal branches were identified in initial gene family trees. Gene families were inferred using IQ-TREE (*90*) using the best-fit model as determined by the Bayesian Information Criterion in each case. Photosynthesis-related COGs were analyzed separately to ensure comprehensive representation of photosystem genes. For more details, see SM section 2.

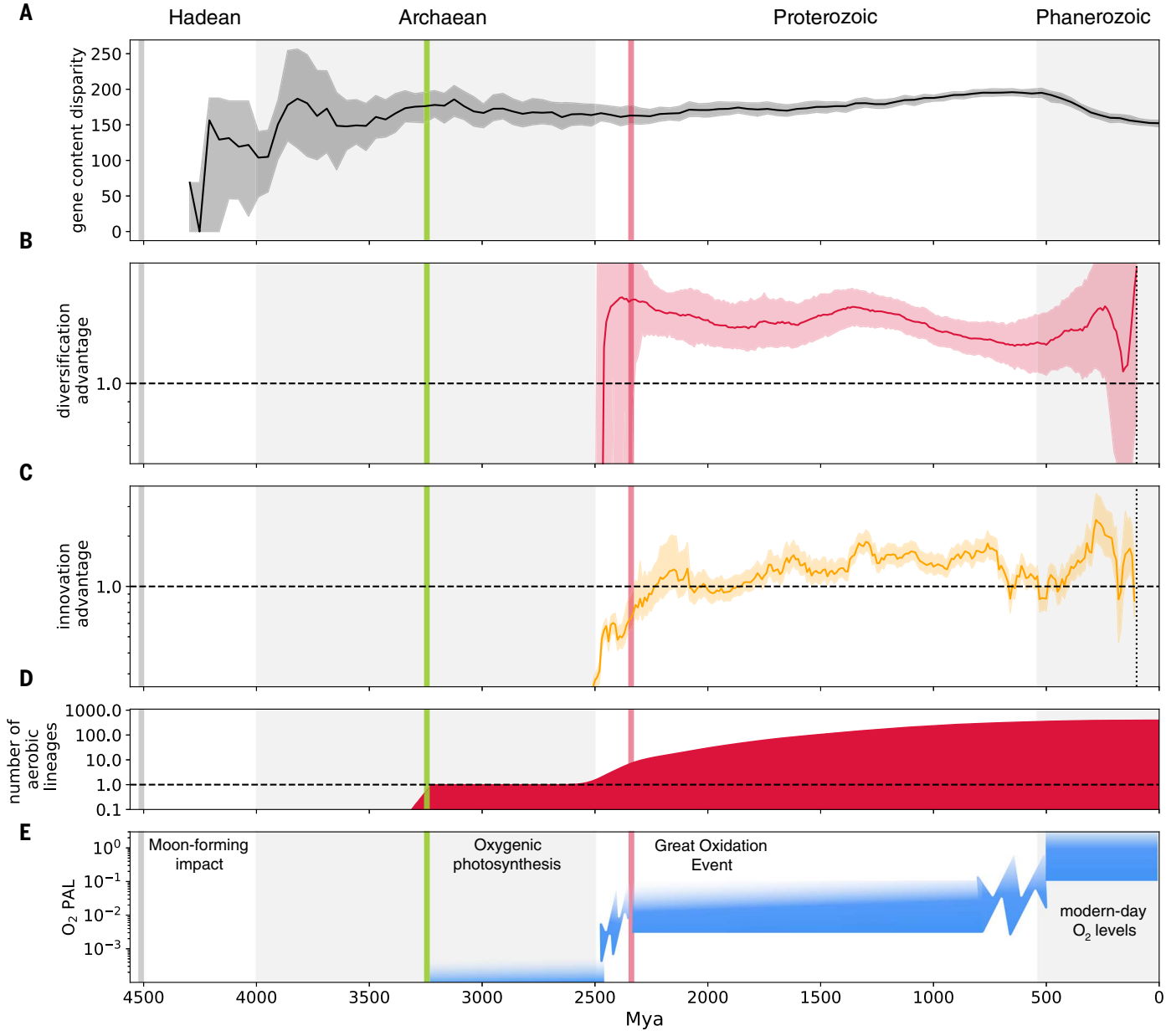


Fig. 6. Bacterial evolution through geological time. (**A**) The early evolution of bacterial genetic diversity in the Archaean (*83*) is reflected by a rapid increase in gene content disparity (see SM section 7). (**B**) The line charts the relative diversification rate (the ratio of the inverse of the waiting time to the next speciation event) for aerobic and anaerobic lineages, revealing a persistent higher diversification rate of aerobes for most of Earth's history since the GOE. Rates equilibrated prior to the end of the Proterozoic, before a return to higher diversification rates for aerobes since the GOE. The dotted line denotes a ratio of 1, where the diversification rate of aerobic and anaerobic lineages is equal. Terminal branches were omitted and ratios since 100 Ma were not calculated as a result of a low number of nodes. (**C**) Spread of aerobic metabolism by HGT. Transitions from anaerobic to aerobic lifestyles have outnumbered transitions in the reverse direction for most of Earth's history since the GOE. The plots show the ratio of aerobic to anaerobic transition rates on the dated species tree during each time period. The dotted line denotes a ratio of 1, where transitions occur in both directions at the same rate. The small number of nodes since ~200 Ma greatly increases uncertainty for both analyses, as reflected in the wide confidence intervals of posterior node age estimates. Ratios since 100 Ma were not calculated due to a low number of nodes. (**D**) Number of aerobic lineages according to Fig. 3. (**E**) Qualitative sketch of atmospheric oxygen levels through time based on (*11*, *13*, *84*). The gray line shows the MFI at ~4.52 Ga; the green line at ~3.23 Ga reflects the presence of fossil and isotopic evidence for oxygenic photosynthesis; the red line at ~2.33 Ga reflects the end of the GOE (see calibrations in SM section 4). Alternating background shading distinguishes geological eons.

Prediction of oxygen use phenotypes

To map anaerobe-to-aerobe transitions in the bacterial tree, we trained a series of machine learning classifiers to predict oxygen use phenotype (aerobe or anaerobe, based on classifications from BacDive) (*91*) from COG gene family counts. The XGBoost method had the best overall accuracy and robustness. The model was trained on extant genomes and then applied to ancestral genomes. For more details, see SM section 3.

Molecular dating analyses

Bayesian relaxed molecular clock analyses were performed with mcmc-date using the calibrations described in SM section 4. We explored a range of analysis conditions to evaluate the robustness

of our results to different calibration approaches (for example, with and without GOE soft maxima) and alternative interpretations of the geological record (SM section 5). The resulting dated species trees were compared using a new time-constrained reconciliation model, implemented in ALE and described in SM section 6.

REFERENCES AND NOTES

1. E. Zuckerkandl, L. Pauling, in *Evolving Genes and Proteins*, V. Bryson, H. J. Vogel, Eds. (Academic Press, 1965), pp. 97–166.
2. A. J. Drummond, S. Y. W. Ho, M. J. Phillips, A. Rambaut, Relaxed phylogenetics and dating with confidence. *PLOS Biol.* **4**, e88 (2006). doi: 10.1371/journal.pbio.0040088; pmid: 16683862
3. S. Y. W. Ho, S. Duchêne, Molecular-clock methods for estimating evolutionary rates and timescales. *Mol. Ecol.* **23**, 5947–5965 (2014). doi: 10.1111/mec.12953; pmid: 25290107
4. M. dos Reis, P. C. J. Donoghue, Z. Yang, Bayesian molecular clock dating of species divergences in the genomics era. *Nat. Rev. Genet.* **17**, 71–80 (2016). doi: 10.1038/nrg.2015.8; pmid: 26688196
5. W. K. Hartmann, D. R. Davis, Satellite-sized planetesimals and lunar origin. *Icarus* **24**, 504–515 (1975). doi: 10.1016/0019-1035(75)90070-6
6. S. A. Jacobson *et al*., Highly siderophile elements in Earth's mantle as a clock for the Moon-forming impact. *Nature* **508**, 84–87 (2014). doi: 10.1038/nature13172; pmid: 24695310
7. O. Abramov, S. J. Mojzsis, Microbial habitability of the Hadean Earth during the late heavy bombardment. *Nature* **459**, 419–422 (2009). doi: 10.1038/nature08015; pmid: 19458721
8. H. C. Betts *et al*., Integrated genomic and fossil evidence illuminates life's early evolution and eukaryote origin. *Nat. Ecol. Evol.* **2**, 1556–1562 (2018). doi: 10.1038/s41559-018-0644-x; pmid: 30127539
9. A. H. Knoll, K. D. Bergmann, J. V. Strauss, Life: The first two billion years. *Philos. Trans. R. Soc. B.* **371**, 20150493 (2016). doi: 10.1098/rstb.2015.0493; pmid: 27672146
10. T. W. Lyons *et al*., Co-evolution of early Earth environments and microbial life. *Nat. Rev. Microbiol.* **22**, 572–586 (2024). doi: 10.1038/s41579-024-01044-y; pmid: 38811839
11. T. W. Lyons, C. T. Reinhard, N. J. Planavsky, The rise of oxygen in Earth's early ocean and atmosphere. *Nature* **506**, 307–315 (2014). doi: 10.1038/nature13068; pmid: 24553238
12. M. R. Warke *et al*., The Great Oxidation Event preceded a Paleoproterozoic "snowball Earth". *Proc. Natl. Acad. Sci. U.S.A.* **117**, 13314–13320 (2020). doi: 10.1073/pnas.2003090117; pmid: 32482849
13. S. W. Poulton *et al*., A 200-million-year delay in permanent atmospheric oxygenation. *Nature* **592**, 232–236 (2021). doi: 10.1038/s41586-021-03393-7; pmid: 33782617
14. N. G. Carr, B. A. Whitton, *The Biology of Cyanobacteria* (Blackwell Scientific Publications, 1982).
15. G. C. Dismukes *et al*., The origin of atmospheric oxygen on Earth: The innovation of oxygenic photosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 2170–2175 (2001). doi: 10.1073/pnas.061514798; pmid: 11226211
16. J. E. Johnson *et al*., Manganese-oxidizing photosynthesis before the rise of cyanobacteria. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 11238–11243 (2013). doi: 10.1073/pnas.1305530110; pmid: 23798417
17. R. M. Soo, J. Hemp, P. Hugenholtz, Evolution of photosynthesis and aerobic respiration in the cyanobacteria. *Free Radic. Biol. Med.* **140**, 200–205 (2019). doi: 10.1016/j.freeradbiomed.2019.03.029; pmid: 30930297
18. W. W. Fischer, J. Hemp, J. S. Valentine, How did life survive Earth's great oxygenation? *Curr. Opin. Chem. Biol.* **31**, 166–178 (2016). doi: 10.1016/j.cbpa.2016.03.013; pmid: 27043270
19. N. J. Planavsky *et al*., Evidence for oxygenic photosynthesis half a billion years before the Great Oxidation Event. *Nat. Geosci.* **7**, 283–286 (2014). doi: 10.1038/ngeo2122
20. J. Jabłońska, D. S. Tawfik, The evolution of oxygen-utilizing enzymes suggests early biosphere oxygenation. *Nat. Ecol. Evol.* **5**, 442–448 (2021). doi: 10.1038/s41559-020-01386-9; pmid: 33633374
21. S. Gribaldo, E. Talla, C. Brochier-Armanet, Evolution of the haem copper oxidases superfamily: A rooting tale. *Trends Biochem. Sci.* **34**, 375–381 (2009). doi: 10.1016/j.tibs.2009.04.002; pmid: 19647436
22. G. P. Fournier *et al*., The Archean origin of oxygenic photosynthesis and extant cyanobacterial lineages. *Proc. Biol. Sci.* **288**, 20210675 (2021). doi: 10.1098/rspb.2021.0675; pmid: 34583585
23. H. Shang, D. H. Rothman, G. P. Fournier, Oxidative metabolisms catalyzed Earth's oxygenation. *Nat. Commun.* **13**, 1328 (2022). doi: 10.1038/s41467-022-28996-0; pmid: 35288554
24. R. L. Morris, T. M. Schmidt, Shallow breathing: Bacterial life at low O(2). *Nat. Rev. Microbiol.* **11**, 205–212 (2013). doi: 10.1038/nrmicro2970; pmid: 23411864
25. D. H. Parks *et al*., A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* **36**, 996–1004 (2018). doi: 10.1038/nbt.4229; pmid: 30148503
26. V. K. Ayyadevara, in *Pro Machine Learning Algorithms: A Hands-On Approach to Implementing Algorithms in Python and R*, V. K. Ayyadevara, Ed. (Apress, 2018), pp. 117–134.
27. G. J. Szöllősi, W. Rosikiewicz, B. Boussau, E. Tannier, V. Daubin, Efficient exploration of the space of reconciled gene trees. *Syst. Biol.* **62**, 901–912 (2013). doi: 10.1093/sysbio/syt054; pmid: 23925510
28. G. J. Szöllősi, E. Tannier, V. Daubin, B. Boussau, The inference of gene trees with species trees. *Syst. Biol.* **64**, e42–e62 (2015). doi: 10.1093/sysbio/syu048; pmid: 25070970
29. J. Martijn, J. Vosseberg, L. Guy, P. Offre, T. J. G. Ettema, Deep mitochondrial origin outside the sampled alphaproteobacteria. *Nature* **557**, 101–105 (2018). doi: 10.1038/s41586-018-0059-5; pmid: 29695865
30. S. A. Muñoz-Gómez *et al*., Site-and-branch-heterogeneous analyses of an expanded dataset favour mitochondria as sister to known Alphaproteobacteria. *Nat. Ecol. Evol.* **6**, 253–262 (2022). doi: 10.1038/s41559-021-01638-2; pmid: 35027725
31. R. I. Ponce-Toledo *et al*., An Early-Branching Freshwater Cyanobacterium at the Origin of Plastids. *Curr. Biol.* **27**, 386–391 (2017). doi: 10.1016/j.cub.2016.11.056; pmid: 28132810
32. P. S. Garcia, F. Barras, S. Gribaldo, Components of iron-Sulfur cluster assembly machineries are robust phylogenetic markers to trace the origin of mitochondria and plastids. *PLOS Biol.* **21**, e3002374 (2023). doi: 10.1371/journal.pbio.3002374; pmid: 37939146
33. P. M. Shih, N. J. Matzke, Primary endosymbiosis events date to the later Proterozoic with cross-calibrated phylogenetic dating of duplicated ATPase proteins. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 12355–12360 (2013). doi: 10.1073/pnas.1305813110; pmid: 23776247
34. T. A. Mahendrarajah *et al*., ATP synthase evolution on a cross-braced dated tree of life. *Nat. Commun.* **14**, 7456 (2023). doi: 10.1038/s41467-023-42924-w; pmid: 37978174
35. C. Brochier-Armanet, E. Talla, S. Gribaldo, The multiple evolutionary histories of dioxygen reductases: Implications for the origin and evolution of aerobic respiration. *Mol. Biol. Evol.* **26**, 285–297 (2009). doi: 10.1093/molbev/msn246; pmid: 18974088
36. M. Wang *et al*., A universal molecular clock of protein folds and its power in tracing the early history of aerobic metabolism and planet oxygenation. *Mol. Biol. Evol.* **28**, 567–582 (2011). doi: 10.1093/molbev/msq232; pmid: 20805191
37. A. M. Satkoski, N. J. Beukes, W. Li, B. L. Beard, C. M. Johnson, A redox-stratified ocean 3.2 billion years ago. *Earth Planet. Sci. Lett.* **430**, 43–53 (2015). doi: 10.1016/j.epsl.2015.08.007
38. P. C. J. Donoghue, M. J. Benton, Rocks and clocks: Calibrating the Tree of Life using fossils and molecules. *Trends Ecol. Evol.* **22**, 424–431 (2007). doi: 10.1016/j.tree.2007.05.005; pmid: 17573149
39. V. Daubin, G. J. Szöllősi, Horizontal Gene Transfer and the History of Life. *Cold Spring Harb. Perspect. Biol.* **8**, a018036 (2016). doi: 10.1101/cshperspect.a018036; pmid: 26801681
40. A. A. Davín *et al*., Gene transfers can date the tree of life. *Nat. Ecol. Evol.* **2**, 904–909 (2018). doi: 10.1038/s41559-018-0525-3; pmid: 29610471
41. C. Magnabosco, K. R. Moore, J. M. Wolfe, G. P. Fournier, Dating phototrophic microbial lineages with reticulate gene histories. *Geobiology* **16**, 179–189 (2018). doi: 10.1111/gbi.12273; pmid: 29384268
42. J. M. Wolfe, G. P. Fournier, Horizontal gene transfer constrains the timing of methanogen evolution. *Nat. Ecol. Evol.* **2**, 897–903 (2018). doi: 10.1038/s41559-018-0513-7; pmid: 29610466
43. H. Shimodaira, M. Hasegawa, CONSEL: For assessing the confidence of phylogenetic tree selection. *Bioinformatics* **17**, 1246–1247 (2001). doi: 10.1093/bioinformatics/17.12.1246; pmid: 11751242
44. G. J. Szöllosi, B. Boussau, S. S. Abby, E. Tannier, V. Daubin, Phylogenetic modeling of lateral gene transfer reconstructs the pattern and relative timing of speciations. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 17513–17518 (2012). doi: 10.1073/pnas.1202997109; pmid: 23043116
45. B. J. Harris *et al*., Rooting Species Trees Using Gene Tree-Species Tree Reconciliation. *Methods Mol. Biol.* **2569**, 189–211 (2022). doi: 10.1007/978-1-0716-2691-7_9; pmid: 36083449
46. W. F. Bottke, M. D. Norman, The Late Heavy Bombardment. *Annu. Rev. Earth Planet. Sci.* **45**, 619–647 (2017). doi: 10.1146/annurev-earth-063016-020131
47. T. A. Mahendrarajah *et al*., ATP synthase evolution on a cross-braced dated tree of life. *Nat. Commun.* **14**, 7456 (2023). doi: 10.1038/s41467-023-42924-w; pmid: 37978174
48. J. S. Boden, J. Zhong, R. E. Anderson, E. E. Stüeken, Timing the evolution of phosphorus-cycling enzymes through geological time using phylogenomics. *Nat. Commun.* **15**, 3703 (2024). doi: 10.1038/s41467-024-47914-0; pmid: 38697988
49. F. U. Battistuzzi, S. B. Hedges, A major clade of prokaryotes with ancient adaptations to life on land. *Mol. Biol. Evol.* **26**, 335–343 (2009). doi: 10.1093/molbev/msn247; pmid: 18988685
50. E. R. R. Moody *et al*., An estimate of the deepest branches of the tree of life from ancient vertically evolving genes. *eLife* **11**, e66695 (2022). doi: 10.7554/eLife.66695; pmid: 35190025
51. K. Mateos *et al*., The evolution and spread of sulfur cycling enzymes reflect the redox state of the early Earth. *Sci. Adv.* **9**, eade4847 (2023). doi: 10.1126/sciadv.ade4847; pmid: 37418533
52. C. Parsons, E. E. Stüeken, C. J. Rosen, K. Mateos, R. E. Anderson, Radiation of nitrogen-metabolizing enzymes across the tree of life tracks environmental transitions in Earth history. *Geobiology* **19**, 18–34 (2021). doi: 10.1111/gbi.12419; pmid: 33108025
53. P. Sánchez-Baracaldo, G. Bianchini, J. D. Wilson, A. H. Knoll, Cyanobacteria and biogeochemical cycles through Earth history. *Trends Microbiol.* **30**, 143–157 (2022). doi: 10.1016/j.tim.2021.05.008; pmid: 34229911
54. Y. Li *et al*., A genome-scale phylogeny of the kingdom Fungi. *Curr. Biol.* **31**, 1653–1665.e5 (2021). doi: 10.1016/j.cub.2021.01.074; pmid: 33607033
55. J. C. Avise, G. C. Johns, Proposal for a standardized temporal scheme of biological classification for extant species. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 7358–7363 (1999). doi: 10.1073/pnas.96.13.7358; pmid: 10377419
56. J. C. Avise, J.-X. Liu, On the temporal inconsistencies of Linnean taxonomic ranks. *Biol. J. Linn.* **102**, 707–714 (2011). doi: 10.1111/j.1095-8312.2011.01624.x
57. M. dos Reis *et al*., Uncertainty in the Timing of Origin of Animals and the Limits of Precision in Molecular Timescales. *Curr. Biol.* **25**, 2939–2950 (2015). doi: 10.1016/j.cub.2015.09.066; pmid: 26603774
58. M. Dohrmann, G. Wörheide, Dating early animal evolution using phylogenomic data. *Sci. Rep.* **7**, 3599 (2017). doi: 10.1038/s41598-017-03791-w; pmid: 28620233
59. B. J. Harris *et al*., Divergent evolutionary trajectories of bryophytes and tracheophytes from a complex common ancestor of land plants. *Nat. Ecol. Evol.* **6**, 1634–1643 (2022). doi: 10.1038/s41559-022-01885-x; pmid: 36175544
60. M. Göker, A. Oren, Proposal to include the categories kingdom and domain in the International Code of Nomenclature of Prokaryotes. *Int. J. Syst. Evol. Microbiol.* **73**, ijsem.0.005650 (2023). doi: 10.1099/ijsem.0.005650; pmid: 36749690
61. R. M. Soo, J. Hemp, D. H. Parks, W. W. Fischer, P. Hugenholtz, On the origins of oxygenic photosynthesis and aerobic respiration in Cyanobacteria. *Science* **355**, 1436–1440 (2017). doi: 10.1126/science.aal3794; pmid: 28360330
62. J. W. Schopf, Precambrian Paleobiology: Problems and Perspectives. *Annu. Rev. Earth Planet. Sci.* **3**, 213–249 (1975). doi: 10.1146/annurev.ea.03.050175.001241
63. R. M. Schwartz, M. O. Dayhoff, Origins of prokaryotes, eukaryotes, mitochondria, and chloroplasts. *Science* **199**, 395–403 (1978). doi: 10.1126/science.202030; pmid: 202030
64. K. M. Towe, Aerobic respiration in the Archaean? *Nature* **348**, 54–56 (1990). doi: 10.1038/348054a0; pmid: 11536471
65. J. S. Berg *et al*., How low can they go? Aerobic respiration by microorganisms under apparent anoxia. *FEMS Microbiol. Rev.* **46**, fuac006 (2022). doi: 10.1093/femsre/fuac006; pmid: 35094062
66. B. Pastina, J. A. LaVerne, Hydrogen Peroxide Production in the Radiolysis of Water with Heavy Ions. *J. Phys. Chem. A* **103**, 1592–1597 (1999). doi: 10.1021/jp984433o
67. H. He *et al*., An abiotic source of Archean hydrogen peroxide and oxygen that pre-dates oxygenic photosynthesis. *Nat.*

Commun. **12**, 6611 (2021). doi: 10.1038/s41467-021-26916-2; pmid: 34785682
68. J. Stone, J. O. Edgar, J. A. Gould, J. Telling, Tectonically-driven oxidant production in the hot biosphere. *Nat. Commun.* **13**, 4529 (2022). doi: 10.1038/s41467-022-32129-y; pmid: 35941147
69. X. Wu *et al*., Geodynamic oxidation of Archean terrestrial surfaces. *Commun. Earth Environ.* **4**, 1–9 (2023). doi: 10.1038/s43247-023-00789-3
70. S. L. Olson, L. R. Kump, J. F. Kasting, Quantifying the areal extent and dissolved oxygen concentrations of Archean oxygen oases. *Chem. Geol.* **362**, 35–43 (2013). doi: 10.1016/j.chemgeo.2013.08.012
71. S. V. Lalonde, K. O. Konhauser, Benthic perspective on Earth's oldest evidence for oxygenic photosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 995–1000 (2015). doi: 10.1073/pnas.1415718112; pmid: 25583484
72. D. Y. Sumner, I. Hawes, T. J. Mackey, A. D. Jungblut, P. T. Doran, Antarctic microbial mats: A modern analog for Archean lacustrine oxygen oases. *Geology* **43**, 887–890 (2015). doi: 10.1130/G36966.1
73. A. M. Nicolas *et al*., Soil Candidate Phyla Radiation Bacteria Encode Components of Aerobic Metabolism and Co-occur with Nanoarchaea in the Rare Biosphere of Rhizosphere Grassland Communities. *mSystems* **6**, e0120520 (2021). doi: 10.1128/msystems.01205-20; pmid: 34402646
74. J. N. Hamm *et al*., Unexpected host dependency of Antarctic Nanohaloarchaeota. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 14661–14670 (2019). doi: 10.1073/pnas.1905179116; pmid: 31253704
75. V. La Cono *et al*., Symbiosis between nanohaloarchaeon and haloarchaeon is based on utilization of different polysaccharides. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 20223–20234 (2020). doi: 10.1073/pnas.2007232117; pmid: 32759215
76. H. Santos *et al*., Aerobic metabolism of carbon reserves by the "obligate anaerobe" Desulfovibrio gigas. *Biochem. Biophys. Res. Commun.* **195**, 551–557 (1993). doi: 10.1006/bbrc.1993.2081; pmid: 8373395
77. A. Das, R. Silaghi-Dumitrescu, L. G. Ljungdahl, D. M. Kurtz Jr., Cytochrome bd oxidase, oxidative stress, and dioxygen tolerance of the strictly anaerobic bacterium Moorella thermoacetica. *J. Bacteriol.* **187**, 2020–2029 (2005). doi: 10.1128/JB.187.6.2020-2029.2005; pmid: 15743950
78. H. I. Boga, W. Ludwig, A. Brune, Sporomusa aerivorans sp. nov., an oxygen-reducing homoacetogenic bacterium from the gut of a soil-feeding termite. *Int. J. Syst. Evol. Microbiol.* **53**, 1397–1404 (2003). doi: 10.1099/ijs.0.02534-0; pmid: 13130024
79. W. W. Fischer, J. Hemp, J. E. Johnson, Evolution of Oxygenic Photosynthesis. *Annu. Rev. Earth Planet. Sci.* **44**, 647–683 (2016). doi: 10.1146/annurev-earth-060313-054810
80. T. A. Williams *et al*., Phylogenetic reconciliation: Making the most of genomes to understand microbial ecology and evolution. *ISME J.* **18**, wrae129 (2024). doi: 10.1093/ismejo/wrae129; pmid: 39001714
81. T. Oliver *et al*., The evolution and evolvability of photosystem II. *Annu. Rev. Plant Biol.* **74**, 225–257 (2023). doi: 10.1146/annurev-arplant-070522-062509; pmid: 36889003
82. A. Nishihara, Y. Tsukatani, C. Azai, M. K. Nobu, Illuminating the coevolution of photosynthesis and Bacteria. *Proc. Natl. Acad. Sci. U.S.A.* **121**, e2322120121 (2024). doi: 10.1073/pnas.2322120121; pmid: 38875151
83. L. A. David, E. J. Alm, Rapid evolutionary innovation during an Archaean genetic expansion. *Nature* **469**, 93–96 (2011). doi: 10.1038/nature09649; pmid: 21170026
84. A. J. Krause, B. J. W. Mills, A. S. Merdith, T. M. Lenton, S. W. Poulton, Extreme variability in atmospheric oxygen levels in the late Precambrian. *Sci. Adv.* **8**, eabm8191 (2022). doi: 10.1126/sciadv.abm8191; pmid: 36240275
85. R. G. Stockey *et al*., Sustained increases in atmospheric oxygen and marine productivity in the Neoproterozoic and Palaeozoic eras. *Nat. Geosci.* **17**, 667–674 (2024). doi: 10.1038/s41561-024-01479-1
86. T. W. Lyons, C. W. Diamond, N. J. Planavsky, C. T. Reinhard, C. Li, Oxygenation, Life, and the Planetary System during Earth's Middle History: An Overview. *Astrobiology* **21**, 906–923 (2021). doi: 10.1089/ast.2020.2418; pmid: 34314605
87. C. T. Reinhard, N. J. Planavsky, The History of Ocean Oxygenation. *Annu. Rev. Mar. Sci.* **14**, 331–353 (2022). doi: 10.1146/annurev-marine-031721-104005; pmid: 34416124
88. J. Pérez, J. Muñoz-Dorado, T. de la Rubia, J. Martínez, Biodegradation and biological treatments of cellulose, hemicellulose and lignin: An overview. *Int. Microbiol.* **5**, 53–63 (2002). doi: 10.1007/s10123-002-0062-3; pmid: 12180781
89. M. Groussin *et al*., Unraveling the processes shaping mammalian gut microbiomes over evolutionary time. *Nat. Commun.* **8**, 14319 (2017). doi: 10.1038/ncomms14319; pmid: 28230052
90. B. Q. Minh *et al*., IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020). doi: 10.1093/molbev/msaa015; pmid: 32011700
91. L. C. Reimer *et al*., BacDive in 2019: Bacterial phenotypic data for High-throughput biodiversity analysis. *Nucleic Acids Res.* **47**, D631–D636 (2019). doi: 10.1093/nar/gky879; pmid: 30256983
92. G. J. Szollosi, T. Williams, A. Davín, B. J. Woodcroft, Data and code for "A geological timescale for bacterial evolution and oxygen adaptation," Version 5, Figshare (2024); https://doi.org/10.6084/M9.FIGSHARE.23899299.
93. H. Shimodaira, An approximately unbiased test of phylogenetic tree selection. *Syst. Biol.* **51**, 492–508 (2002). doi: 10.1080/10635150290069913; pmid: 12079646
94. P. Sánchez-Baracaldo, T. Cardona, On the origin of oxygenic photosynthesis and Cyanobacteria. *New Phytol.* **225**, 1440–1446 (2020). doi: 10.1111/nph.16249; pmid: 31598981
95. T. Oliver, P. Sánchez-Baracaldo, A. W. Larkum, A. W. Rutherford, T. Cardona, Time-resolved comparative molecular evolution of oxygenic photosynthesis. *Biochim. Biophys. Acta Bioenerg.* **1862**, 148400 (2021). doi: 10.1016/j.bbabio.2021.148400; pmid: 33617856

ACKNOWLEDGMENTS

We thank J. A. Palacios for providing R scripts to compute the D2 distance for ranked trees. We are grateful for the help and support provided by the Scientific Computing and Data Analysis section of Core Facilities at OIST. **Funding:** This work was funded by the following: European Research Council (ERC) grant 714774 "GENECLOCKS" (to A.A.D., L.L.S., D.S., and G.J.S.) under the European Union's Horizon 2020 research and innovation program; ERC grant 947317" ASymbEL" (to A.S.) under the European Union's Horizon 2020 research and innovation program; Gordon and Betty Moore Foundation grant GBMF9741 (to T.A.W., A.S., P.C.J.D., and G.J.S.); Gordon and Betty Moore Foundation's Symbiosis in Aquatic Systems Initiative grant GBMF9346 (to A.S.); Moore–Simons Project on the Origin of the Eukaryotic Cell, Simons Foundation 735929LPI grant 735929LP (to A.S.); Royal Society University Research Fellowship (to T.A.W.); John Templeton Foundation grant 62220 (to E.R.R.M., D.P., P.C.J.D., and T.A.W.); Leverhulme Trust Research Fellowship grant RF-2022-167 (to P.C.J.D.); Biotechnology and Biological Sciences Research Council grants BB/T012773/1 and BB/Y003624/1 (to P.C.J.D.); Australian Research Council (ARC) Future Fellowship grant FT210100521 (to B.J.W.); ARC Discovery Project grant DP230101171 (to B.J.W.); ARC Discovery Early Career Researcher Award grant DE190100008 (to R.M.S.); ARC Laureate Fellowship grant FL150100038 (to A.A.D. and P.H.); ARC Discovery Project grant DP220100900 (to A.A.D. and P.H.). **Author contributions:** Conceptualization: A.A.D., A.S., B.B., B.J.W., D.P., E.R.R.M., G.J.S., J.H., P.C.J.D., P.H., R.M., R.M.S., T.A.W., and W.W.F. Method development and coding: A.A.D., B.B., B.J.W., B.M., D.S., G.J.S., L.L.S., and T.A.W. Taxon selection and genomic data preparation: A.A.D., A.S., P.H., and T.A.W. Phylogenetic analyses: A.A.D., B.B., E.R.R.M., G.J.S., L.L.S., and T.A.W. Fossil Calibrations: D.P., P.C.J.D., and T.A.W. Molecular dating: D.S., G.J.S., L.L.S., and S.Á.-C. Machine Learning: B.B., B.J.W., and G.J.S. HCO curation and analysis: R.M., R.M.S., and J.H. Visualization: A.A.D., A.S., B.J.W., E.R., G.J.S., J.W.C., and T.A.W. Writing: A.A.D., A.S., B.B., B.J.W., B.M., D.P., D.S., E.R., E.R.R.M., G.J.S., J.W.C., J.H., L.L.S., P.C.J.D., P.H., R.M., R.M.S., S.Á.-C., T.A.W., and W.W.F. **Competing interests:** R.M. is currently affiliated with Dept Life Sciences, University of Nevada, Las Vegas. All other authors declare no other competing interests. **Data and materials availability:** All data used are described in detail in the SM and available in the figshare repository associated with the submission (*92*). All code used is described in detail in the SM and available either openly on github or other relevant public repository, as indicated, or included in the figshare repository (*92*) associated with the submission in the case of more specialized scripts.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adp1853
Materials and Methods
Supplementary Text
Figs. S1 to S39
Tables S1 to S8
References (*96*–*221*)
MDAR Reproducibility Checklist

Submitted 12 March 2024; accepted 19 December 2024
10.1126/science.adp1853

A geological timescale for bacterial evolution and oxygen adaptation

Adrián A. Davín, Ben J. Woodcroft, Rochelle M. Soo, Benoit Morel, Ranjani Murali, Dominik Schrempf, James W. Clark, Sandra Álvarez-Carretero, Bastien Boussau, Edmund R. R. Moody, Lénárd L. Szánthó, Etienne Richy, Davide Pisani, James Hemp, Woodward W. Fischer, Philip C. J. Donoghue, Anja Spang, Philip Hugenholtz, Tom A. Williams, and Gergely J. Szöll#si

Science **388** (6742), eadp1853. DOI: 10.1126/science.adp1853

Editor's summary

When exploring deep time, the problem is that there are few, if any, good fossils of the earliest living organisms, and it is impossible to precisely date the evolution of those that do exist. One calibration point is provided by the impact event about 4.5 billion years ago that resulted in sterilization of Earth and formation of the Moon. Davín *et al*. used molecular clocks, machine learning, and phylogenetic reconciliation to present a reconstruction of the evolution of Earth's bacterial biosphere over the past 4 billion years with particular emphasis on aerobic metabolisms. Their analysis showed that the last common ancestor of bacteria likely existed 4.4 to 3.9 billion years ago, and aerobic organisms likely emerged before the Great Oxidation Event (2.43 to 2.33 billion years ago). Oxygen tolerance may have been a prerequisite for, rather than a consequence of, the evolution of oxygenic photosynthesis. —Caroline Ash