Timing the evolution of antioxidant enzymes in cyanobacteria

Joanne S. Boden1, Kurt O. Konhauser2, Leslie J. Robbins3,4 & Patricia Sánchez-Baracaldo1✉

The ancestors of cyanobacteria generated Earth’s first biogenic molecular oxygen, but how they dealt with oxidative stress remains unconstrained. Here we investigate when superoxide dismutase enzymes (SODs) capable of removing superoxide free radicals evolved and estimate when Cyanobacteria originated. Our Bayesian molecular clocks, calibrated with microfossils, predict that stem Cyanobacteria arose 3300–3600 million years ago. Shortly afterwards, we find phylogenetic evidence that ancestral cyanobacteria used SODs with copper and zinc cofactors (CuZnSOD) during the Archaean. By the Paleoproterozoic, they became genetically capable of using iron, nickel, and manganese as cofactors (FeSOD, NiSOD, and MnSOD respectively). The evolution of NiSOD is particularly intriguing because it corresponds with cyanobacteria’s invasion of the open ocean. Our analyses of metalloenzymes dealing with reactive oxygen species (ROS) now demonstrate that marine geochemical records alone may not predict patterns of metal usage by phototrophs from freshwater and terrestrial habitats.
Oxygen is essential for complex life forms as it is used during aerobic respiration to create more energy per mol of substrate than other available electron acceptors. Although today the Earth’s atmosphere contains ~21% oxygen (O₂), it was at least 10³ times lower in the Archean 4.0–2.5 billion years ago (Gya)²,³. Just how and when O₂ first appeared as a byproduct of biological evolution-oxic photosynthesis - remains controversial, with estimates ranging from near the origin of life⁴ to 3.8 billion years ago (Ga)⁵ to immediately preceding the Great Oxidation Event (GOE)⁶, which is estimated to have begun 2.50–2.45 Ga⁷,⁸. Before this, O₂ could have been produced by physical processes, such as photodissociation of water and carbon dioxide by UV light⁹,¹⁰, but is unlikely to have accumulated at appreciable levels. With the evolution of oxygenic photosynthesis came the potential to produce O₂ on a much larger scale. Since O₂ is highly reactive, early Cyanobacteria—the first producers of biogenic O₂—likely experienced selective pressure, resulting in the evolution of more efficient antioxidants. Such effects have been documented in the photosynthetic machine, which has been evolving strategies of dealing with reactive oxygen species (ROS) throughout its history¹¹. Therefore, it is reasonable to assume that O₂-generating organisms, such as cyanobacteria, would have co-evolved more efficient mechanisms of managing ROS as water oxidation proteins evolved.

Cyanobacteria remove ROS using carotenoids, α-tocopherol, and antioxidant enzymes, including peroxydases, catalases, superoxide reductases (SORs), and superoxide dismutases (SODs)¹², and their evolutionary history can be elucidated with phylogenetic methodologies. Although peroxydases and catalases enhance the rate of ROS removal (such as H₂O₂ and R-O-OH)¹³, SORs and SODs remove superoxide free radicals (O₂⁻)¹⁴. These O₂⁻ are produced as a byproduct of photosynthetic and respiratory electron transport chains¹² as well as extracellular processes on the cell surface¹⁵. They can also have beneficial roles in iron acquisition, cell signaling, and growth¹⁵,¹⁶, but if O₂⁻ are allowed to accumulate inside the cell, they react with solvent-exposed 4Fe-4S clusters in proteins, including those required for amino acid biosynthesis¹⁷ and photosynthesis,¹⁸ generating reactants of the Fenton reaction, which can ultimately lead to extensive DNA damage¹². To balance the beneficial effects of O₂⁻ with damage caused by over-exposure, organisms must maintain control over their abundance. Perhaps, it is for this reason that SODs and SORs have been found in all three domains of life—Eukarya, Archaea, and Bacteria¹⁴.

Phylogenetic methodologies coupled with protein structural analyses have revealed that three different isoforms of SOD enzymes evolved independently of one another to remove unnecessary O₂⁻. Each has a unique 3D structure, amino acid sequence, and metal cofactor(s); either manganese (MnSOD), nickel (NiSOD), or a combination of copper and zinc (CuZnSOD)¹⁹,²⁰. A fourth SOD enzyme shares its evolutionary heritage with MnSOD, but utilizes iron as its cofactor (FeSOD).²¹ All of these SODs are found within cyanobacteria²²-²³, but their locations within the cell differ. FeSOD and NiSOD are cytoplasmic,²²,²⁴-²⁷ whereas CuZnSODs from Synechococcus and MnSODs from Anabaena and Plectolyngbya are tethered to membranes²⁵-²⁷. The choice of which SOD isoform(s) and associated metal cofactor(s) are used by a given species could reflect its evolutionary heritage, function, and environmental history.

The evolutionary record of cyanobacteria spans at least 1.88 Gya (the oldest undisputed colonial cyanobacterial microfossils),³¹, and perhaps as many as or >3.32 Gya (based on stromatolite fabric³² and molecular clock analyses of PSI³³). During this time, cyanobacteria have diversified into a wide range of marine, freshwater, and terrestrial habitats.³⁴,³⁵ Early phylogenetic studies on the evolution of SODs in cyanobacteria were limited by the availability of genomes but found that NiSODs are encoded in the genomes of planktonic marine strains (such as Prochlorococcus spp.), whereas MnSOD and FeSOD are more widespread²²,²³,³⁶. CuZnSODs, while rare, are not restricted to any particular group²²,²³. The 3D structures required for metalloproteins to incorporate copper cofactors have previously been predicted to have arisen after those required to incorporate Fe- and Mn-utilizing proteins in the Proterozoic (~2–0.5 Gya), following the GOE.²⁷ Therefore, it would follow that cyanobacteria used Fe-SODs and MnSODs before CuZnSODs. Although genes encoding CuZnSODs and FeSODs/MnSODs are widely distributed in bacteria and eukaryotes, NiSODs are restricted to bacteria and may have appeared later in the evolutionary history of life.³⁸ Previous approaches aimed at determining metal usage in SODs have not considered the habitats where strains live or evolved. This is problematic because metal availability differs across oceanic and terrestrial environments³⁸,³⁹.

Results

Bacterial SOD diversity. To elucidate which transition metals cyanobacteria first used to protect themselves against the oxidative stress caused by O₂⁻, the evolutionary history of SOD genes was modeled and mapped onto an updated time-calibrated phylogeny. In order to do this, we began by screening 15,899 bacterial genomes for genes encoding NiSOD, CuZnSOD, and MnSOD. BLASTP was unable to distinguish between genes encoding FeSOD and MnSOD, so they are considered as one hereafter. Results reveal that the sodN gene encoding NiSOD is less common in bacteria than sodC encoding CuZnSOD. The sodN gene was found in 1,464 different species, including gammaproteobacteria, alphaproteobacteria, planctomycetes, bacteroidetes, actinobacteria, cyanobacteria, verrucomicrobia, and chloroflexi (Fig. 1a). By contrast, sodC was absent from the latter two phyla, but present in more genomes (5,723) of beta-, gamma- and alpha–proteobacteria as well as planctomycetes, bacteroidetes, actinobacteria, cyanobacteria, and firmicutes (Fig. 1b). Together, sodA and sodB are more widespread than sodC and sodN combined, being present in 13,748 bacterial genomes across all ten phyla mentioned previously, as well as fibrobacteres, chlorobi, and vamiprovibirionia (Fig. 1c and Supplementary Fig. 1).

Cyanobacterial SOD diversity. A more specific search among cyanobacteria revealed that most strains with sodN (51 of 55 strains) live in saltwater habitats (Fig. 2). They include representatives from all major clades of marine taxa (Fig. 3). Ten lack a gene (named sodX) encoding NiSOD’s maturation protease (Supplementary Data 1). This maturation protease activates NiSOD by cleaving the preprotein,⁴³ so these strains may not be able to use NiSOD to remove O₂⁻. Six of them contain genes encoding other SOD isoforms, while the remaining four pico-cyanobacteria (i.e., Prochlorococcus and Synechococcus species) do not (Fig. 3). The genome completeness of free-living strains—
Fig. 1 Evolutionary relationships of superoxide dismutases in bacteria. a NiSOD b CuZnSOD c Fe- and Mn-utilizing SODs. a and b are representative of three independent replicates (Supplementary Figs. 11–12), whereas c is the most likely of three potential topologies (Supplementary Fig. 1). All trees were constructed with ML methodology implemented in IQ-TREE v1.6.146. Branches are colored to represent proteins found in 1 of 13 bacterial phyla: Cyanobacteria (orange), actinobacteria (green), gammaproteobacteria (blue), alphaproteobacteria (yellow), betaproteobacteria (pink), bacteroidetes (cyan), firmicutes (red), chloroflexi (light green), verrucomicrobia (dark blue), planctomycetes (purple), chlamydia (olive), fibrobacteres (dark green), and chlorobi (gray). Branch lengths represent the number of amino-acid substitutions per site with scale bars representing an average of one substitution per site. Numbered arrows indicate cyanobacterial proteins. Interesting UFBoot values ≥ 95 in all NiSOD replicate trees are indicated with black circles.

Fig. 2 Habitat distribution of SOD genes. Panels represent the distribution of CuZnSOD (red), NiSOD (orange), and Fe- and Mn-utilizing SODs (blue) in cyanobacteria from a marine and b non-marine habitats. Non-marine habitats include c freshwater, d geothermal springs, and e terrestrial habitats.
Fig. 3 Time-calibrated cyanobacterial tree of life and superoxide dismutases. Genes encoding NiSOD (orange circles), NiSOD maturation protease (orange stars), CuZnSOD (red squares), and SODs with Mn- or Fe- cofactors (blue triangles) are highlighted next to the strain names. The earliest node predicted to have sodN, sodC and sodA or sodB are annotated with labels (see Table 2 for posterior age probabilities and Supplementary Fig. 10 for the age distribution of these events). The phylogenetic tree was estimated from SSU and LSU ribosomal RNA and 136 core cyanobacterial proteins from 167 different taxa using the maximum likelihood methodology implemented in IQ-TREE v1.6.146. Node labels represent ultrafast bootstrap approximations <100. Ages were estimated using a Bayesian relaxed molecular clock with Uncorrelated Gamma Multipliers49 for ribosomal RNA. Black circles represent calibration points (Table 1). The first divergence of cyanobacteria was constrained to occur between 2.7 Ga and 2.32 Ga110,111. The color behind each strain name indicates whether it is from a marine habitat (light blue), freshwater habitat (gray), terrestrial habitat (light brown), or geothermal spring (light pink). GOE Great Oxidation Event, Uni. diaz. unicellular diazotrophs, Phan. Phanerozoic eon, V. Vampirovibrio.
measured with BUSCO v3.0.2 using lineage data from cyanobacteria is estimated at 92–98% so it is possible that the maturation protease gene is present but has not been sequenced (Supplementary Table 1).

Many cyanobacterial strains contain multiple SOD isoforms. These include 32 strains with sodN, every strain with sodC, and 45 of 115 with sodA or sodB. The remaining 23 strains which only use NiSOD (including three without sodX) are all marine with small genomes. They include picocyanobacteria and endosymbionts in larger marine tunicates, algae, and sponges (e.g., Prochloron spp., UCYNA, and two strains of Synechococcus spongianum). The 70 cyanobacteria which only use FeSOD and MnSOD live in a variety of marine, terrestrial, and freshwater habitats (Fig. 3). Only 3% of cyanobacteria (five of 149 strains discounting plastids) have genes encoding every SOD isoform as well as the NISOD maturation protease (Fig. 3). Paralogues of sodC were found in three genomes and paralogues of sodA and/or sodB were found in at least 13 genomes (Supplementary Data 1).

The resources needed to make each SOD isoform vary. NiSOD is composed of 157 amino acids (range 145–166), whereas CuZnSOD is composed of 177 (range 103–236) and FeSOD/MnSOD of 200 (range 197–280) (Supplementary Fig. 2).

**Transfer of SODs across phyla.** To investigate whether cyanobacteria obtained their SOD genes from other bacterial phyla, maximum likelihood (ML) phylogenies were constructed using the NiSODs, CuZnSODs, and combined Mn- and Fe-SODs from bacteria. If, in each case, all cyanobacterial proteins form a monophyletic group (i.e., they share a recent common ancestor), this would indicate a single origin. Surprisingly, these phylogenetic analyses revealed that cyanobacterial NiSODs, CuZnSODs, MnSODs, and FeSODs are polyphyletic (Fig. 1 and Supplementary Fig. 1), suggesting independent origins perhaps due to several lateral gene transfer events or other modes of reticulated evolution. Furthermore, the SODs with Mn or Fe cofactors found in vampirovibrioibians are not related to those in their sister phyla, cyanobacteria (Fig. 1c, Supplementary Table 2 and Supplementary Fig. 1).

Molecular phylogenies identified several horizontal gene transfers (HGTs) of each SOD isoform between cyanobacteria and other bacterial phyla: Two of sodN (NiSODs), five of sodC, and three to eight of Mn- and Fe-utilizing SODs (Fig. 1 and Supplementary Fig. 1). A variety of different phyla are involved in these HGTs. For example, most cyanobacterial NiSODs have diversified from a protein resembling that of benthic marine Deltaproteobacteria (namely *Geopsychrobacter electrodiphilus*, UFBoot 88 and *Plesiocystis pacifica*, UFBoot 89, Supplementary Fig. 3), whereas other cyanobacterial NiSODs (present in *Synechococcus spongianum*) diversified from a protein resembling those of alphaproteobacteria (UFBoot 81, Supplementary Fig. 4).

Genes encoding CuZnSOD may have been present in the shared common ancestor of all extant cyanobacteria. This ancestor gave rise to basal lineages before diverging into macrocyanobacteria and microcyanobacteria (Fig. 3). Although CuZnSODs are rare in macro- and micro–cyanobacteria, they are present in most free-living basal lineages (Fig. 3). Two in particular (*Pseudanabaena* spp. and *Gloeobacter* spp.) share sodC genes which are monophyletic (PP 1) and closely related in the same way as the species are to one another (Supplementary Fig. 5). This suggests they have been vertically inherited from the common ancestor of all extant cyanobacteria (otherwise known as the cyanobacteria crown group).

As cyanobacteria diversified to occupy new ecological niches and habitats, the *sodC* genes which initially allowed crown cyanobacteria to use copper and zinc to protect against oxidative stress were likely lost. Later, HGTs may have occurred between non-cyanobacterial phyla and picocyanobacteria, diazotrophs, sponge symbionts, *Acaryochloris* spp., and *Microcoleus chthonoplastes* (Fig. 1), resulting in the distribution found today.

**Bayesian relaxed molecular clock analyses.** Divergence times were estimated in Phylolabes 4.1 using SSU and LSU ribosomal RNA from 164 cyanobacteria and eight vampirovibrioibria. The molecular clock’s topology was constrained using an ML tree constructed in IQ-TREE v1.6.1 from these same ribosomal RNAs plus 136 core cyanobacterial proteins with similar evolutionary trajectories. They have a range of functions in metabolism, cellular processes, and information handling. All molecular clock analyses implemented six calibration points (Table 1) as previously described. Divergence times were estimated using uncorrelated gamma multipliers. All analyses, regardless of calibration points and models, indicate that cyanobacteria diverged from vampirovibrioibria between 3.3 and 3.6 Ga in the Archeaneon (Table 2). Estimates that include 95% credibility intervals allow for a range between 2.8 and 4.3 Ga, suggesting that cyanobacteria diverged from their sister phyla at at least 300 million years before the GOE (Table 2).

**Table 1** Calibration points used for molecular clock analyses.

| Calibration                          | Minimum age/mya | References | Maximum age /mya | References | Diversification between...
|--------------------------------------|-----------------|------------|------------------|------------|-------------------------|
| 1st cyanobacterial diversification  | 2320            | 116        | 2700             | 56         | Gloeobacter violaceus PCC 7421 |
| Filamentous cyanobacteria           | 2320            | 110        | 3000             | 56         | Acaryochloris sp. MBIC 11017 |
| Akinete-forming cyanobacteria       | 1900            | 103        | n/a              | n/a        | Pseudanabaena biceps PCC 7429 |
| Apical cells of cyanobacteria       | 1600            | 105        | 1888             | 102,104    | Leptolyngbya sp. PCC 7376 |
| Endosymbionts of *Hemiaulus* spp.   | 1700            | 106        | 1888             | 102,104    | Calothrix sp. PCC 7103 |
| Endosymbionts of *B. begelowii*     | 110             | 108        | n/a              | n/a        | Nostoc azollae 0708 |

Note: Genus and species names are italicized.
### Table 2 Statistics summarizing divergence times predicted by molecular clock analyses.

| First divergence of cyanobacteria* | 3.0–3.2 Ga | 2.7–3.2 Ga |
|-----------------------------------|------------|------------|
| Divergence of cyanobacteria       | 3544       | 3374       |
| from vampirovibrioia              | (4235–3001) | (4058–2807) |
| NiSOD                             | 912        | 806        |
| (1952–281)                       | (1898–226) |
| CuZnSOD                           | 2926       | 2649       |
| (3043–2692)                      | (2720–2490) |
| MnSOD/FeSOD                       | 1140       | 1045       |
| (1941–207)                      | (1893–171) |

Mean divergence times are presented in millions of years alongside 95% confidence intervals in brackets.

* describes the calibration strategy used to constrain the first radiation of cyanobacteria, where Gloeobacter spp. diversify from other cyanobacteria.

### Discussion

The advent of oxygenic photosynthesis had profound impacts on Earth’s climate, chemistry, and biota. It destroyed a methane greenhouse which kept the world warm\(^50\), enhanced the supply of iron formations (BIF)\(^54\). Organisms would have responded by chemical oxidation of dissolved Fe(II) in seawater to form banded Shih, et al.\(^55\) are calibrated with non-cyanobacterial fossils. Our variations in taxa selection, and alternative molecular markers of age estimates is owing to the choice of molecular clock model, var-

The Neoarchaean (\(\sim 3.8 \text{ Ga}\)) represents the first major transition to a new oxygenated world, where cyanobacteria (Fig. 3) emerge from the Archaean, at least 2.9 \(\text{ Ga}\)\(^55\). If proteins could incorporate Ni into MnSODs in the Archaean, at least 2.9 \(\text{ Ga}\) (Table 2). This fits with alternative phylogenetic analyses predicting that sodC was present in the last universal common ancestor of life on Earth\(^65\).

Our conclusion is based on the close relationship of CuZnSODs from two basal lineages: *Pseudanabaena* spp. and *Gloeobacter* spp. (Supplementary Fig. 5). The most recent common ancestor (MRCA) of these two genera appeared \(\sim 3.4 \text{ Ga}\) (confidence intervals span 4.2–2.8 Ga), and diversified \(\sim 2.6 \text{ Ga}\) (confidence intervals span 3.0 to 2.5 Ga), giving rise to all extant cyanobacteria (Fig. 3). If the ancestor of crown cyanobacteria had sodC, it could have been inherited by its descendants, resulting in a unique phylogenetic signal whereby the CuZnSODs of *Gloeobacter* spp. and *Pseudanabaena* spp. are sisters. Because this phylogenetic pattern exists (Supplementary Fig. 5), it is likely that Archaean cyanobacteria used CuZnSOD. An alternative scenario could result in the same pattern if sodC was transferred laterally from the MRCA of *Pseudanabaena* spp. to the MRCA of *Gloeobacter* spp. This would put an upper age constraint on the origin of CuZnSOD \(\sim 0.95 \text{ Ga}\) (confidence intervals span 2.2 to 0.3 Ga) (Fig. 3).

Although sodC has been transferred between bacterial phyla on multiple occasions in the past (Fig. 1b), our phylogenetic analyses provide less evidence of HGT between cyanobacteria of terrestrial and freshwater origin (see Supplementary Discussion). As the MRCA of *Pseudanabaena* and *Gloeobacter* spp. were not marine (Supplementary Fig. 6), HGT accompanied by a more recent origin of CuZnSOD \(\sim 0.95 \text{ Ga}\) seems unlikely, if not impossible.

Further insight can be gained by timing the origin of Cu-based metalloenzymes in general. Characteristic protein-folds required to bind Cu are predicted to have evolved during, or following, the GOE\(^37\). However, geological evidence suggests that metabolisms which rely on copper metalloenzymes (e.g., nitrification), were present in the Archaean, \(2.7 \text{ Ga}\)\(^64\). If proteins could incorporate Cu cofactors in the Archaean, then CuZnSODs could have been present at the root of cyanobacteria as a mechanism to deal with \(\text{O}_2\) generated by photosynthesis before it accumulated in the global atmosphere (at the GOE). This would have been particularly important for Mesoarchaean and Neoarchaean individuals living inside benthic mats\(^65,66\), where there was a smaller diffusion gradient to pull \(\text{O}_2\) out of the cells. As a result, higher concentrations of \(\text{O}_2\) accumulated, raising the potential for superoxide radical generation, particularly in the periplasm\(^57\). In the alternative (if unlikely) scenario where basal cyanobacteria acquire sodC via HGT \(\sim 0.95 \text{ Gya}\), Archaean cyanobacteria might not have used SOD to remove \(\text{O}_2\) because we do not find any phylogenetic evidence for NiSOD, MnSOD, or FeSOD having been inherited from an ancestor present before the GOE (Fig. 4).

Two scenarios that could reconcile an earlier origin of NiSOD and Fe- or Mn-utilizing SODs with our evolutionary trees (Supplementary Figs. 7–8) are presented in the Supplementary Discussion. All rely on more extensive HGT, so age distributions are presented in Fig. 4, but in paler colors, because most prokaryotic proteins are transferred via vertical inheritance\(^68–70\).

Why are there only a few cyanobacterial strains with the gene encoding CuZnSOD today? Perhaps it is related to physiological considerations resulting in the replacement of CuZnSOD for MnSODs. The thylakoid membranes of *Nostoc* sp. PCC7120 lack CuZnSOD (Fig. 3) and instead contain MnSOD\(^30\). When exposed to intense sunlight these MnSODs protect the cell from photoinhibition\(^60\) in a similar manner to CuZnSOD\(^25\). Furthermore, the gene encoding MnSOD can be post-translationally processed to produce three proteins of different sizes which vary in concentration between heterocysts and vegetative cells\(^29\). Under nitrogen-supplemented conditions, the smallest cytosolic 24 kDa protein predominates the slightly larger 27 kDa protein,
but under nitrogen-limiting conditions, both proteins are present in near equal proportions. No such flexibility has been documented in CuZnSODs. Therefore, the ability to modify the size and localization of MnSOD proteins in response to changes in the environment could explain why the sodA gene encoding MnSOD is present in all major clades of nitrogen-fixing cyanobacteria (defined by , including all Nostocales, one Trichodesmium spp., seven unicellular diazotrophs, Chroococcidiopsis thermalis, and Leptolyngbya sp. PCC7375, Fig. 3) and many phyla of non-photosynthetic bacteria with diazotrophic representatives (e.g., Firmicutes, Actinobacteria and Proteobacteria). Phylogenetic evidence for the utilization of SODs that incorporate Fe and Mn cofactors appears relatively recently, in the middle of the Proterozoic (Table 2, Fig. 4). Intriguingly, modern cyanobacteria might be less sensitive to oxidative stress caused by O$_2$ because their proteins with 4Fe-4S clusters have evolved to protect themselves by positioning their Fe atoms in less-solvent-exposed parts of the molecule. It may be with this method that picocyanobacteria are able to thrive with only a cytoplasmic NiSOD (Fig. 3).

The relatively late appearance of FeSOD and MnSOD in cyanobacteria is surprising as previous studies have postulated an Archean origin in bacteria. What caused this delay? Earth system models suggest that photoferrotrophs outcompeted cyanobacteria for upwelling nutrients in aquatic habitats prior to the GOE. Perhaps, they also limited the soluble Fe$^{2+}$ available for cyanobacteria, thus facilitating a selective advantage to lineages that used alternative metals for relieving oxidative stress. As the atmosphere became more oxygenated, photoferrotrophs were marginalized to shrinking pools of Fe$^{2+}$. Our analyses suggest that the Neoproterozoic oxygenation increased cyanobacteria’s requirement for ROS defense mechanisms so much that lineages began using FeSOD or MnSOD regardless of the waning global concentrations of Fe and Mn (Fig. 4). Further study, however, will be needed to assess how effectively FeSODs, MnSODs, and CuZnSODs protect against oxidative stress under different ambient concentrations of oxygen and transition metals.

Phylogenetic evidence also indicates that cyanobacteria started using cytosolic NiSOD isoforms in the MRCA of Leptolyngbya sp. PCC7375 and Nodosilinea nodulosa PCC7104, which diversified in marine Neoproterozoic (Table 2, Fig. 4) habitats (Supplementary Fig. 1). Phylogenetic evidence suggests that cyanobacteria started using CuZnSODs at the end of the Precambrian. Although only speculative at this time, it is possible that the acquisition of NiSOD and its associated maturation protease assisted the invasion of cyanobacteria into pelagic marine habitats. Pelagic planktonic cyanobacteria are typically limited by P, Fe, and N (in cases of non-nitrogen fixers) owing to their distance from sources of riverine discharge. As NiSOD is composed of fewer amino acids than other SOD isoforms (Supplementary Fig. 2) and does not require Fe, its utilization by benthic marine cyanobacteria in the Neoproterozoic may have imparted an evolutionary advantage and increased their resilience to nutrient limitation in the open ocean.

By studying only extant taxa, it is inherently difficult to estimate whether extinct lineages of cyanobacteria used NiSOD,
MnSOD, FeSOD, or CuZnSOD prior to the estimations provided in Fig. 3, Fig. 4, and Table 2. It is reassuring, however, that our estimates of all SOD isoforms predate the Cretaceous–Paleogene mass extinction of non-avian dinosaurs from terrestrial and aquatic environments 66 Mya\(^8\), the Permian-Triassic mass extinction of 56% of all marine animal genera 252 Mya\(^9\), and the Snowball Earth glaciations 720–635 Mya, which subjected life to extreme climate fluctuations\(^82\). Therefore, our estimations of timing are robust across several past extinction events. Any future discovery of novel lineages of basal cyanobacteria, however, could alter our estimated order of SOD appearance.

It has been proposed that trace metal inventories of ancient marine sediments should reflect trace metal availability and correspond to the emergence of novel metabolisms and metalloenzymes\(^27\). In light of this suggestion, the past two decades have seen a number of papers purporting the use of ancient marine chemical sediments (i.e., BIF, shales, pyrite) to track seawater composition through time (see ref. \(^{79}\) for a review). With regards to the metal cofactors contained in SOD, their temporal trajectories have been reconstructed based on both thermodynamic considerations (e.g.\(^83\)) and sedimentary records (e.g.\(^{84,85}\)). However, there is a poor record of trace metal inventories outside of marine habitats, so it can be more difficult to establish links between metal availability and diversification of cyanobacteria living in terrestrial and freshwater habitats.

Our ancestral state reconstructions predict that a non-marine cyanobacterium could have acquired CuZnSOD (Supplementary Fig. 6) in the Archaean (Table 2). In the Archaean to Proterozoic ocean, Cu and Zn were supposed to have been present at exceedingly low abundances given the low solubility of sulfoxide in the water column, which subjected life to extreme climate fluctuations\(^82\). Therefore, our estimations of timing are robust across several past extinction events. Any future discovery of novel lineages of basal cyanobacteria, however, could alter our estimated order of SOD appearance.

Given the histories of these trace metals, one may ask whether it is availability driving the assimilation of Cu, Zn, Fe, Mn, and Ni into SOD metalloenzymes, or an inherent competition with other extant lineages at any given time. For instance, metalloenzyme fold superfamilies that incorporate Zn proliferate late and have been proposed to contribute to the delay in eukaryotic evolution\(^37\). Alternatively, this has been proposed to reflect an intrinsic biological property of eukaryotic evolution rather than a sudden shift in Zn availability\(^84\). In the latter view, the early incorporation of Cu and Zn into SODs would reflect a lack of competition from the yet unevolved eukaryotes in the Archaean. Then in the late Mesoproterozoic, as eukaryotes began to become more abundant and dominate marine and terrestrial environments, they may have outcompeted cyanobacteria for Cu and Zn, providing an impetus for the emergence of other membrane-bound SODs, such as MnSOD—effectively providing a strategy for alleviating limitation imparted by competition with emergent eukaryotes. Although eukaryotes also utilize Mn, protein structures required to bind Mn have not been preferentially retained in their genomes\(^84\) and cyanobacteria, having evolved earlier than photosynthetic eukaryotes\(^95\), had plenty of time to develop an efficient Mn uptake system.

Later, in Neoproterozoic marine habitats, cyanobacteria began supplementing their ROS defense mechanisms with a new and smaller SOD isoform (NiSOD) that requires less phosphorus and nitrogen to manufacture. This timing of sodN incorporation into the cyanobacterial genomic repertoire coincides with Earth’s second planetary oxygenation\(^3\) and the emergence of planktonic marine cyanobacteria\(^35\). Therefore, we suggest that cyanobacteria began assimilating nickel into SODs to invade Neoproterozoic planktonic communities as Neoproterozoic oxygen levels rose.

**Methods**

**Acquisition of SOD sequences.** Bacterial SOD diversity was assessed by downloading 15,899 bacterial genomes from the NCBI RefSeq database (https://www.ncbi.nlm.nih.gov/refseq/) in 2018. They were searched for genes encoding NiSOD, CuZnSOD, FeSOD, and MnSOD using the basic local alignment search tool (BLAST) for proteins (https://blast.ncbi.nlm.nih.gov/Blast.cgi) with a word size of six, gap opening penalty of 11, gap extension penalty of 1, and e-value cut off of \(1 \times 10^{-5}\). Query sequences are detailed in Supplementary Table 3. They were aligned to relevant subject sequences using the BLOSUM62 substitution matrix. Each genome was treated as a separate subject (functionally equivalent to a unique domain) for the NCBI conserved domain search (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) (Supplementary Data 3 and Supplementary Fig. 9), if possible, query sequences with experimentally defined functions were utilized.

Cyanobacterial SOD diversity was assessed by sampling 155 additional genomes (strains are listed in Supplementary Data 1 and their habitat in Supplementary Table 4), including representatives from all major clades of salt-tolerant and freshwater cyanobacteria\(^18\). These were supplemented with five eukaryotic plastid genomes\(^96\) and eight vamproibionial genomes representing all four orders (Gastranaerophilales, Vampirovibrionales, Caenarcaniphilales, and Obscuribacteriales)\(^90\).
SOD gene phylogenies. Amino-acid sequences for all SOD isoforms (including paralogs found in cyanobacteria) were aligned using MAFFT96 and gaps present in ≥80% of sequences removed. The evolutionarily related history of each isoform was estimated by constructing nine ML phylogenies with IQ-TREE version 1.6.146. Substitution models were identified by ModelFinder97 and node supports measured using ultrafast bootstrap two approximations98.

Species phylogeny of cyanobacteria. An evolutionary tree of cyanobacteria was generated from a data set including SSU rRNA (1687 nucleotides), LSU rRNA (3387 nucleotides), and 136 core proteins (52,227 aa). Each protein is encoded by an orthogonal gene, which is conserved among cyanobacteria and involved in a key cellular function, such as information processing, metabolism, or cellular processes. For a full description of these genes, see previous papers34,35,77. Each protein and ribosomalosomal RNA were aligned using MAFFT96 and concatenated using an alignment viewer (http://sdsssdld.altervista.org/arklumpus/AlignmentViewer/AlignmentViewer.html). Gaps were removed if present at the same position in an alignment viewer (http://sdsssdld.altervista.org/arklumpus/AlignmentViewer/AlignmentViewer.html). Gaps were removed if present at the same position in an alignment viewer.

Divergence time estimation. Bayesian relaxed molecular clocks were implemented using the topology described above. This topology was fixed, and ages estimated based on predicted mutation rates for ribosomal RNA (SSU and LSU, 5074 aligned positions). Substitutions were modeled using a flexible general time-reversible model inferred from the alignment and a Dirichlet process prior99 (CAT-GTR) to account for different rates of evolution between distant sites of the molecule. Divergence times were estimated in phylobayes 4.110 using uncorrelated gamma multipliers,49 a birth-death prior on divergence times, and root prior chosen from a Gamma probability distribution with mean 3060 and standard deviation 404. As a result, 97% of the prior distribution specified that the ancestor of all vampirovibrio and cyanobacteria originated after the end of the late heavy bombardment 3.9 Ga100,101.

We also implemented microfossil calibrations as follows; filamentous cyanobacteria >1.9 Ga102,103, akinete-forming cyanobacteria 1.6–1.888 Ga102,104,105, and apical cells of cyanobacteria 1.7–1.888 Ga102,104,105. These fossil constraints were supplemented with evidence dating the appearance of eukaryotic hosts of endosymbiotic cyanobacteria. For example, Richelia species diversify before their diatom host, named Hemisphaera107, appeared 110 Mya109, and UCYN-A species diversify before their prymnesiophyte host, named Braarudosphaera bigelowii, appeared 91 Mya109. Geochemical evidence was used to constrain cyanobacteria to first diversify before the Great Oxygenation Event of 2.32 Ga110, but after either; (a) 3 Ga when molybdenum oxides document the first “whiffs” of atmospheric oxygen46; or (b) 2.7 Ga based on the earliest fossilized evidence of cyanobacterial stromatolites111. Soft bounds were applied throughout to allow 5% of the prior probability density to fall outside of the specified minimum and maximum ages.

Models were considered complete when four replicate independent chains converged. This was tested by estimating effective sample sizes and relative differences using Tracemc (in Phylobybes with effsize > 50, reldiff < 0.3). Chronograms and divergence times were calculated from a single representative chain using Readdiv (in Phylobybes) with the same burn-in (25% of mean chain length) and sampling frequency (1 in every 10 points) as Traceomp.

Chronology of SOD isoforms. The order in which SOD isoforms appeared in cyanobacteria, was estimated using a method described as “topological comparison”112. First, a Bayesian protein phylogeny was created for each SOD isoform using only cyanobacterial sequences (Supplementary Figs. 5 and 7). Bayesian phylogenetic reconstructions of MnSODs and FeSODs had not converged after 2 weeks, so the sequences were separated into seven groups based on their position in the ML bacterial phylogenies (Supplementary Fig. 1) to conduct alignments and phylogenetic reconstructions on more similar sequences (Supplementary Fig. 8). The resulting protein phylogenies were then compared with the species phylogeny of cyanobacteria (Fig. 3) to identify monophyletic groups whose NiSOD, CuZnSOD, or MnSOD/FeSOD had evolved as expected by vertical inheritance (without HGT). The last common ancestor of each of these clades was assumed to have been capable of using the corresponding SOD.

Ancestral state reconstruction of habitat preference. To find out which habitats ancestral cyanobacteria lived in when SOD isoforms appeared, Bayesian stochastic character mapping113 was implemented with SIMMAP v1.5114 in the phytools package115 of R using our time-calibrated trees. Prior distribution on the root node of the tree was estimated based on the data, 1000 simulations were performed and the all rates different model was utilized to allow transition rates between marine and non-marine habitats to vary based on the data (as implemented in95). Character states were coded as either “marine” or “non-marine” (Supplementary Data 1).

Compilation of Ni data. From a database of >4000 literature shale analyses spanning the Archaean to modern, 1584 had both Ni and Ti data available and are used here to reconstruct the trajectory of Ni in seawater (Supplementary Data 4). Molar Ni/Ti ratios were normalized to evolving continental crust116 then time-binned based on age (Archaean, Paleoproterozoic, Mesoproterozoic, Neoproterozoic, Phanerozoic). This normalization accounts for secular changes in the composition of Earth’s continental crust that reflect the emergence, growth, and subsequent differentiation of continental crust from the Archaean through the Phanerozoic. The evolution of continental crust influences the relative availability of trace and major elements weathered from land to the oceans, and thus the normalization removes this temporal variability, facilitating a comparison of Ni/Ti ratios in shales through time. Mean values for each time bin were bootstrap resampled (n = 10,000; Fig. 5) in Matlab* 2019b using the bootstrap function.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability. The sequence data analyzed in this study and accession numbers of genomes from the NCBI RefSeq database (https://www.ncbi.nlm.nih.gov/refseq/) are available in the open science framework repository, https://osf.io/7y7q/?view_only=cc61929817e4913a7795cde10ae64f.
2. Kump, L. R. The rise of atmospheric oxygen. Nature 451, 277–278 (2008).
3. Lyons, T. W., Reinhard, C. T. & Planavsky, N. J. The rise of oxygen in Earth's early ocean and atmosphere. Nature 506, 307–315 (2014).
4. Oliver, T., Sanchez-Baracaldo, P., Larkum, A. W., Rutherford, A. W. & Cardona, T. Time-resolved comparative molecular evolution of oxygenic photosynthesis. Biochim. Biophys. Acta Bioenerg. 1862, 148400 (2021).
5. Revsbech, N. P. & Frie, R. U-rich Archaea sea-floor sediments from Greenland - indications of >3700 Ma oxygenic photosynthesis. Earth Planet. Sci. Lett. 217, 237–244 (2004).
6. Fischer, W. W., Hemp, J. & Johnson, J. E. Evolution of oxygenic photosynthesis. Annu. Rev. Earth Planet. Sci. 44, 647–683 (2016).
7. Konhauser, K. O. et al. Aerobic bacterial pyrite oxidation and acid rock drainage during the Great Oxidation Event. Nature 478, 369–374 (2011).
8. Warke, M. R. et al. The Great Oxidation Event preceded a Paleoproterozoic "snowball Earth". Proc. Natl Acad. Sci. USA 117, 13314–13320 (2020).
9. Lu, Z., Chang, Y. C., Yin, Q. & Jackson, W. M. Evidence for direct molecular oxygen production in CO2 photosynthesis. Science 346, 61–64 (2014).
10. Meadows, V. S. Reflections on O2 as a biosignature in exoplanetary atmospheres. Astrobiology 17, 1022–1052 (2017).
11. Orf, G. S., Gisriel, C. & Redding, K. E. Evolution of photosynthetic reaction centers: insights from the structure of the holoenzymatic reaction center. Photosynth. Res. 138, 11–37 (2018).
12. Latifi, A., Ruiia, M. & Zhang, C. C. Oxidative stress in cyanobacteria. FEMS Microbiol. Rev. 33, 258–278 (2009).
13. Bernroither, M., Zamocky, M., Furtmuller, P. G., Peschek, G. A. & Obinger, C. Occurrence, phylogeny, structure, and function of catalases and peroxidases in cyanobacteria. J. Exp. Bot. 60, 423–440 (2009).
14. Sheng, Y. et al. Superoxide dismutases and superoxide reductases. Chem. Rev. 118, 8345–8398 (2018).
15. Diaz, J. M. & Plummer, S. Production of extracellular reactive oxygen species by phytoplankton: past and future directions. J. Plankton Res. 40, 655–666 (2018).
16. Rose, A. L. The influence of extracellular superoxide on iron redox chemistry and bioavailability to aquatic microorganisms. Front. Microbiol. 3, 124 (2012).
17. Miller, A. F. Superoxide dismutases: ancient enzymes and new insights. FEBS Lett. 586, 585–595 (2012).
18. Noll, M. et al. The evolution of oxygenic photosynthesis. Science 344, 1467–1471 (2014).
19. Robins, L. et al. Trace elements at the intersection of marine biological and geochemical evolution. Earth-Sci. Rev. 163, 323–348 (2016).
20. Shtil, T. & Falkowski, P. G. Evidence for oxygenic cyanobacteria in 3.5 Ga rocks: implications for the origin of the oxygen atmosphere. Phil. Trans. R. Soc. Lond. B Biol. Sci. 373, 20170132 (2018).
21. Simaan, J. N., You, Y. T., von Haeseler, A. & Minh, B. Q. iQ-TREE: a fast online phylogenetic method for maximum likelihood analysis. Nucleic Acids Res. 44, W232–W235 (2016).
22. Stal, L. J. & Francois, R. Phylogenetic and taxonomic diversity of Prochlorococcus and Synechococcus in the ocean. Appl. Environ. Microbiol. 74, 6549–6556 (2008).
23. Summons, S. A., Lefebvre, P. E., Retherford, S. L. & Lough, G. M. The early evolution of photosynthetic bacteria belonging to a new candidate phylum siphonata to a new candidate phylum. Science 344, 1467–1471 (2014).
24. Bashan, Y. & Dekel, A. M. Insights into the evolution of picocyanobacteria and phycoerythrin gene acquisition. J. Phycol. 41, 453–465 (2005).
25. Priya, B. et al. Comparative analysis of cyanobacterial superoxide dismutases to discriminate canonical forms. BMC Genomics 8, 435 (2007).
26. Schad, M., Konhauser, K. O., Sanchez-Baracaldo, P., Kappler, A. & Bryce, C. The evolution of oxygenic photosynthesis in the Archean and its influence on the temporal and spatial development of the microbial iron cycle on ancient Earth? Free Radic. Biol. Med. 140, 156–169 (2020).
27. Shilo, P. M., Hemp, J., Ward, L. M., Matzke, N. J. & Fischer, W. W. Cyanobacterial superoxide dismutase interacts with iron and copper redox cycling in cyanobacterial superoxide dismutase.
58. Louca, S. et al. Bacterial diversification through geological time. Nat. Ecol. Evol. 2, 1458–1467 (2018).
59. Jablonowski, C. & Tawfik, D. S. The evolution of oxygen-utilizing enzymes suggests early biosphere oxygenation. Nat. Ecol. Evol. 5, 442–448 (2021).
60. Zhao, W. X., Guo, Q. X. & Zhao, J. D. A membrane-associated Mn-superoxide dismutase protects the photosynthetic apparatus and nitrogenase from oxidative damage in the cyanobacterium Anabaena sp. PCC 7120. Plant Cell Physiol. 48, 563–572 (2007).
61. Sae-Tang, P. et al. Overexpressed superoxide dismutase and catalase act synergistically to protect the repair of PSII during photoinhibition in Synechococcus elongatus PCC 7942. Plant Cell Physiol. 57, 1899–1907 (2016).
62. Shao, S., Cardona, T. & Nixon, P. J. Early emergence of the FtsH proteases involved in photosystem II repair. Photosynthetica 56, 163–177 (2018).
63. Weiss, M. C. et al. The physiology and habitat of the last universal common ancestor. Nat. Microbiol. 1, 16116 (2016).
64. Moore, E. K., Jelen, B. I., Giovannelli, D., Raanan, H. & Falkowski, P. G. Metal oxides, divalent cations, silica and the early earth phosphorus crisis. Proc. Natl Acad. Sci. USA 113, 6995–1000 (2015).
65. Lalonde, S. V. & Konhauser, K. O. Benthic perspective on Earth transition in the marine nitrogen cycle. Geobiology 70, 225–237 (2013).
66. Wilmeth, D. T. et al. Neoarchean (2.7 Ga) lacustrine stromatolite deposits in the Hartbeesfontein Basin, Ventersdorp Supergroup, South Africa: Implications for cyanobacterial evolution in the cyanobacteria using a compartmentalization approach. Geobiology 3, 145–165 (2005).
67. Hartmann, L. S. & Barnum, S. R. Inferring the evolutionary history of Mo-dependent nitrogen fixation from phylogenetic studies of nifK and nifDK. J. Mol. Evol. 71, 70–85 (2011).
68. Irupakutika, M. A., Sengupta, S., Devireddy, A. R., Azad, R. K. & Mittler, R. The evolution of reactive oxygen species metabolism. J. Exp. Bot. 67, 5933–5946 (2016).
69. Kunin, V., Goldovsky, L., Darzentas, N. & Ouzounis, C. A. The net of life: reconstructing the microbial phylogenetic network. Genome Res 15, 954–959 (2005).
70. Beiko, R. G., Harlow, T. J. & Ragan, M. A. Highways of gene sharing in prokaryotes. Proc. Natl Acad. Sci. USA 102, 14332–14337 (2005).
71. Novichkov, P. S. et al. Genome-wide molecular clock and horizontal gene transfer in bacterial evolution. J. Bacteriol. 186, 6575–6585 (2004).
72. Sanchez-Baracaldo, P., Hayes, P. K. & Blank, C. E. Morphological and habitat evolution in the cyanobacteria using a compartmentalization approach. Geobiology 3, 145–165 (2005).
73. Wilmuth, D. T. et al. Neorcan dualite in the marine nitrogen cycle. Geobiology 70, 225–237 (2013).
74. Hofman, H. J. Precambrian microflora, belcher islands, canada: significance and systematics. J. Paleontol. 50, 1040–1073 (1976).
75. Sergeyev, V. N., Gerasimenko, L. M. & Zavarzin, G. A. The Proterozoic history and present state of cyanobacteria. Microbiology 71, 623–637 (2002).
76. Galubic, S., Sereev, V. N. & Kon, A. H. Mesoproterozoic archeaeilopaidae: akinetes of heterocystous cyanobacteria. Lethonia 28, 285–298 (1995).
77. Zhang, Y. & Galubic, S. Endolithic microfossils (cyanophyta) from early proterozoic stromatolites, heibe, china. Acta Micropalaeontol. Sin. 4, 1–12 (1987).
78. Foster, R. A. et al. Nitrogen fixation and transfer in open ocean diatom-cyanobacterial symbioses. ISME J. 5, 1484–1493 (2011).
79. Sims, A. P., Mann, D. G. & Medlin, L. K. Evolution of the diatoms' insights from fossil, biological and molecular data. Phylogia 545, 361–402 (2006).
80. Cornejo-Castillo, F. M. et al. Cyanobacterial symbionts diverged in the late cretaceous towards lineage-specific nitrogen fixation factories in single-celled phytoplankton. Nat. Commun. 7, 11071 (2016).
81. Bekker, A. et al. Dating the rise of atmospheric oxygen. Geochim. Cosmochim. Acta 68, A780–A780 (2004).
82. Bosak, L., Liang, B., Sim, M. S. & Petroff, A. P. Morphological record of oxygentic photosynthesis in conical stromatolites. Proc. Natl Acad. Sci. USA 106, 10939–10943 (2009).
83. Chan, C. X., Beiko, R. G. & Ragan, M. A. Scaling up the phylogenetic detection of later gene transfer events. Methods Mol. Biol. 1525, 421–432 (2017).
84. Hulsenbeck, J. P., Nielsen, R. & Bollback, J. P. Stochastic mapping of morphological characters. Syst. Biol. 52, 131–158 (2003).
85. Bollback, J. P. SIMMAP: Stochastic character mapping of discrete traits on phylogenies. BMC Bioinformatics 7, 88 (2006).
86. Revell, L. J. Phytools: An R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. 3, 217–223 (2012).
87. Condie, K. C. Chemical-composition and evolution of the upper continental crust - contrasting results from surface samples and shales. Chem. Geol. 104, 1–37 (1993).

Acknowledgements
We thank Andrew H. Knoll, Paul J. Valdes, John Raven, and Paul Falkowski for helpful comments on the manuscript. We also thank Giorgio Bianchini for technical help in preparing the manuscript.
designing scripts for automated BLASTP and leaf coloring as well as developing a trimming tool for alignments. All phylogenetic analyses were performed at high-performance computing facilities (BlueCrystal 3 and 4) at the University of Bristol. Funding support for this work came from a Royal Society University Research Fellowship to P.S.-B. and a University of Bristol Graduate Teaching Scholarship to J.S.B. L.J.R. gratefully acknowledges the support of a Donnelley Environmental Postdoctoral Fellowship from the Yale Institute for Biospheric Studies.

Author contributions
J.S.B. and P.S.-B. conceived the project; J.S.B. and P.S.-B. designed the phylogenetic and molecular clocks analyses, J.S.B. performed analyses, J.S.B. and P.S.-B. interpreted evolutionary results; K.O.K and L.J.R. interpreted the Ni shale records; J.S.B., K.O.K, L.J.R., and P.S.-B. wrote the paper.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-021-24396-y.

Correspondence and requests for materials should be addressed to P.S.-B.

Peer review information Nature Communications thanks Luisa Falcón, Kevin Sutherland, and the other, anonymous, reviewer for their contribution to the peer review of this work.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© Crown 2021