FEBS Letters

FEBS PRESS science publishing by scientists

RESEARCH ARTICLE

The radical impact of oxygen on prokaryotic evolution—enzyme inhibition first, uninhibited essential biosyntheses second, aerobic respiration third

Natalia Mrnjavac[1], Falk S. P. Nagies[1], Jessica L. E. Wimmer[1], Nils Kapust[1], Michael R. Knopp[1], Katharina Trost[1], Luca Modjewski[1], Nico Bremer[1], Marek Mentel[2], Mauro Degli Esposti[3], Itzhak Mizrahi[4], John F. Allen[5] and William F. Martin[1]

1 Institute of Molecular Evolution, Faculty of Mathematics and Natural Sciences, Heinrich Heine University Düsseldorf, Germany
2 Department of Biochemistry, Faculty of Natural Sciences, Comenius University in Bratislava, Bratislava, Slovakia
3 Center for Genomic Sciences, UNAM Campus de Cuernavaca, Mexico
4 Department of Life Sciences, Ben-Gurion University of the Negev and The National Institute for Biotechnology in the Negev, Be'er-Sheva, Israel
5 Research Department of Genetics, Evolution and Environment, University College London, UK

Correspondence
N. Mrnjavac, Institute of Molecular Evolution, Faculty of Mathematics and Natural Sciences, Heinrich Heine University Düsseldorf, Universitätsstrasse 1, 40225 Düsseldorf, Germany
Tel: +49 211 8112736
E-mail: n.mrnjavac@hhu.de

Natalia Mrnjavac, Falk S. P. Nagies, Jessica L. E. Wimmer, and Nils Kapust contributed equally to this article.

(Received 13 February 2024, revised 12 April 2024, accepted 19 April 2024)

doi:10.1002/1873-3468.14906

Edited by Peter Brzezinski

Molecular oxygen is a stable diradical. All O_2-dependent enzymes employ a radical mechanism. Generated by cyanobacteria, O_2 started accumulating on Earth 2.4 billion years ago. Its evolutionary impact is traditionally sought in respiration and energy yield. We mapped 365 O_2-dependent enzymatic reactions of prokaryotes to phylogenies for the corresponding 792 protein families. The main physiological adaptations imparted by O_2-dependent enzymes were not energy conservation, but novel organic substrate oxidations and O_2-dependent, hence O_2-tolerant, alternative pathways for O_2-inhibited reactions. Oxygen-dependent enzymes evolved in ancestrally anaerobic pathways for essential cofactor biosynthesis including NAD^+, pyridoxal, thiamine, ubiquinone, cobalamin, heme, and chlorophyll. These innovations allowed prokaryotes to synthesize essential cofactors in O_2-containing environments, a prerequisite for the later emergence of aerobic respiratory chains.

Keywords: aerobic metabolism; evolution of aerobes; evolution of respiration; great oxidation event; lateral gene transfer; oxygen inhibition

The Great Oxidation Event, GOE [1], divides Earth's history close to its midpoint. Roughly 2.4 billion years ago (Ga), photosynthetic prokaryotes with two chlorophyll-based photosystems linked in series—cyanobacteria—evolved the molecular tools needed to extract electrons from H_2O, and to use them to fix CO_2 and N_2 for growth [2,3]. Oxygen made by cyanobacteria is ground state triplet O_2, a stable diradical with two unpaired electrons of identical spin. Its structure is better written as $^{\bullet}O–O^{\bullet}$ instead of O=O to underscore its diradical nature [4] (Fig. 1). Radicals are molecules having unpaired valence electrons. They have the property of extracting single electrons from available donors so as to restore a stable octet electron configuration, converting the donor into a new radical in the process. Though most radicals are extremely reactive [5], the O_2 diradical is generally unreactive [6]. It is kinetically stable because each of the unpaired electrons in O_2 is delocalized over a two-center, three-electron π bond, resulting in a very

Abbreviations
GOE, great oxidation event; LGT, lateral gene transfer; OEC, oxygen-evolving complex; PAL, present atmospheric level; PLP, pyridoxal phosphate; ROS, reactive oxygen species; SLP, substrate-level phosphorylation; SOD, superoxide dismutase; *V*, verticality.

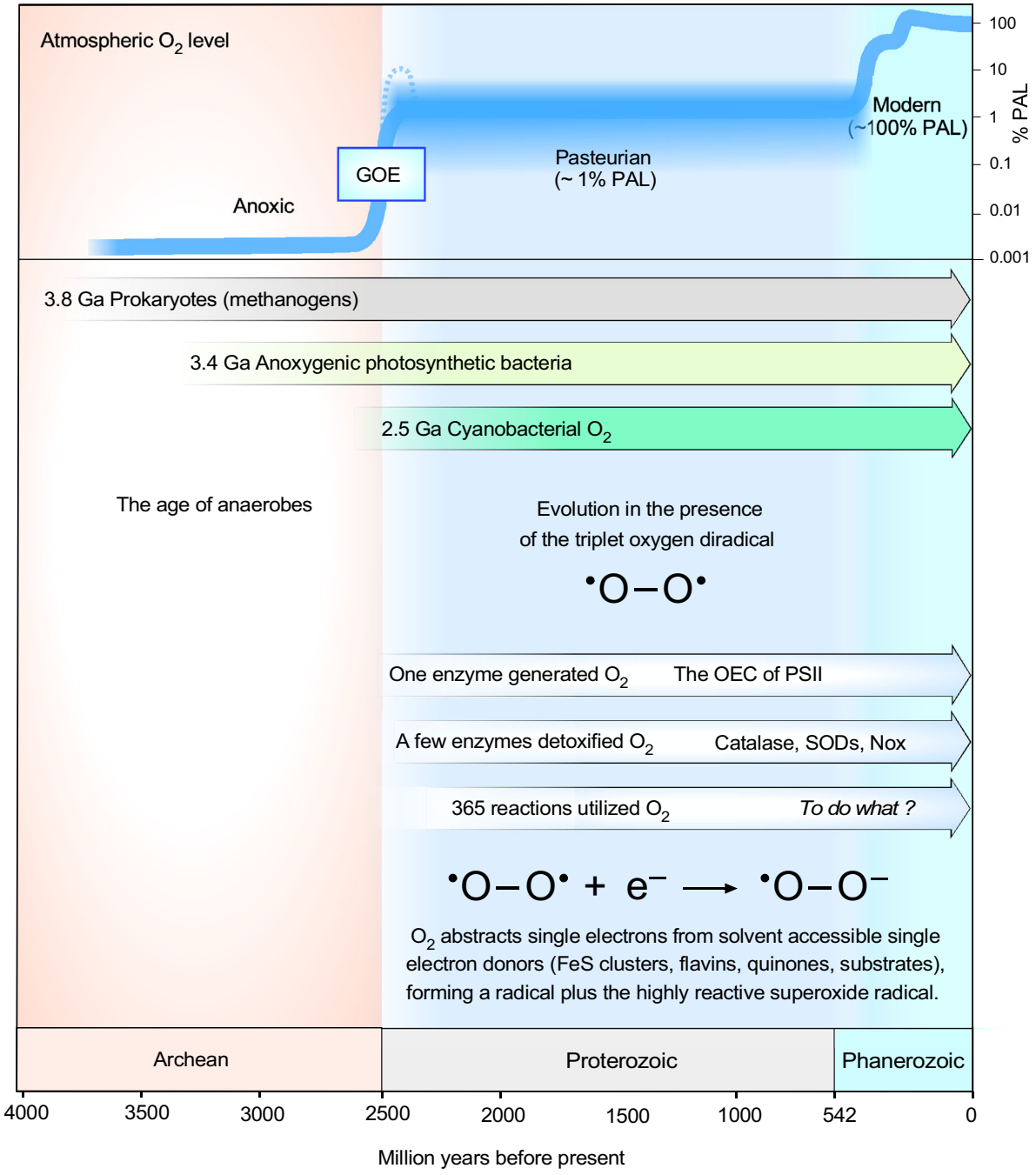


Fig. 1. A timeline of Earth history with the rise of O_2 and the appearance of relevant groups of prokaryotes. Blue boxes emphasize the GOE, the onset of O_2-dependent enzymes, and questions concerning their evolution and functions. The number 365 represents the number of O_2-dependent reactions in KEGG that we mapped to protein families (see text). SOD: superoxide dismutase; Nox: NADH oxidases (diaphorases), which are oxygen detoxifying enzymes [9]. Data from references [1,2,10–14]. Note that ultra-light carbon in 3.8 Ga rocks can be interpreted as evidence of both archaeal methanogens [15] and bacterial acetogens, which carry the same isotope signal [16], because both fix CO_2 *via* the acetyl-CoA pathway. A broken line reaching to 10% PAL around the end of the GOE indicates the Lomagundi excursion [10]. The reasons why O_2 levels remained near the Pasteur point for 1.8 billion years are still discussed. Numerous geological causes [10] and one biological cause [12] for the existence of the boring billion have been proposed. It is undebated that cyanobacteria (and their descendants, plastids) generated the current global supply of O_2 *via* one single enzyme and one single enzyme activity: the conserved Mn_4CaO_5-containing oxygen-evolving complex (OEC) of photosystem II. O_2 is written as •O–O• instead of O=O to underscore its diradical nature [4]. By 3.4–3.3 Ga, anoxygenic photosynthetic prokaryotes were generating stromatolites in aerial settings [17].

large resonance stabilization energy, and consequently a high activation energy barrier [4]. Despite this, O_2 has a very weak σ bond [4], which renders O_2 an extremely energy-rich molecule [7], so energy-rich that it undergoes exergonic redox reactions with every element except gold [8].

From the origin of the first microbes roughly 4 Ga to the onset of the GOE, the Earth's oceans and atmosphere were effectively devoid of O_2 [10] (Fig. 1). At the GOE, O_2 became introduced into Earth's oceans and atmosphere to approximately 1% of its present atmospheric level (PAL), and stayed more or less constant at this level for roughly 1.8 billion years. O_2 levels started to rise again about 580 million years ago, approaching modern values with the advent of land plants ca. 450 million years ago [10–13]. This protracted low oxygen phase of Earth history from 2.4 Ga to 0.58 Ga has been called the 'boring billion' [14] to emphasize the lack of geologically interesting events during that period of O_2 stasis, but it has also been called the 'Pasteurian' era of life's history [18] to emphasize the crucial observation that O_2 levels of 1% PAL correspond to the Pasteur point—the level of ambient oxygen (ca. 1% PAL or 0.2% v/v) at which facultative aerobes switch their terminal acceptors from anaerobic to aerobic respiration. The "boring" Pasteurian billion was the era during which prokaryotes learned to cope with the reactivity of O_2, to use O_2 for their own benefit, and to evolve alternatives to enzymes that were inhibited by O_2 (Fig. 1).

Prokaryotes were eye witnesses to the GOE. Since they lived before, during, and after the GOE (Fig. 1), their O_2-dependent enzymes should hold clues about the impact of O_2 on physiological evolution. To characterize the impact of O_2 in biochemical evolution, we identified 365 enzymatic reactions of prokaryotes that utilize O_2 in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database [19] and that map to phylogenies of prokaryotic genes [20]. The results provide novel insights into the impact that O_2 exerted on microbial evolution, the nature of physiological traits that O_2-dependent enzymes imparted, and the dispersal of genes for O_2-dependent enzymes *via* lateral gene transfer (LGT) following the GOE.

Materials and methods

Collection of oxygen-dependent and oxygen-independent reactions

Data for 11 804 metabolic reactions were downloaded from the KEGG [19] reaction database (version 10th August 2022). Additionally, we manually added the reaction linked to superoxide dismutase (SOD; R00275). In KEGG, SOD was linked to an enzyme commission (EC) number, which was in turn linked to the KEGG orthology identifier (KO) for SOD. However, there was no direct link between the SOD reaction and the KO. As there was no discernable reason for this, we added the enzyme SOD and its reaction for the qualitative descriptions of this paper.

The data of the 11 805 reactions were subsequently filtered for reactions involving O_2 (KEGG compound C00007), yielding a set of 1949 O_2-dependent reactions, most of these being specific to eukaryotes. The reactions were mapped to prokaryotic protein families using KEGG orthology to link reactions to sequences. Protein families were created using MCL [21] as previously described [20,22], and only protein families with 4 or more sequences were used in this analysis. Only reactions that we could link to prokaryotic protein families were retained, yielding 365 O_2-dependent reactions mapped to 792 protein families occurring in prokaryotes, referred to as the O_2-dependent reaction set. Of the remaining reactions, a set of 3018 prokaryotic reactions linked to prokaryotic protein families, making up the O_2-independent reaction set (Table S1). See Data accessibility for access to protein family data. For protein family annotation, all clustered sequences were blasted against the KEGG database using Diamond 2.0.1 [23]. All best hits with at least 25% identity and a maximum *e*-value of 1×10^{-10} were used for annotation. Based on these hits, one KO (KEGG orthology identifier) annotation was assigned to each protein family based on majority rule (Table S2). Protein families which contained at least 75% of unknown sequences without any hits in the KEGG database were not annotated.

Verticality distribution of O_2-dependent and O_2-independent reactions

Verticality values (V) from Nagies *et al.* [20] were assigned to protein families (Table S1). Verticality describes the relative amount of LGT that a gene family has undergone. High verticality indicates little LGT (vertical evolution, typical for ribosomal proteins) whereas low verticality (typical for O_2-dependent enzymes) indicates abundant LGT. Only prokaryotic protein families spanning 2 or more prokaryotic phyla had an associated verticality value [20]. All the phylogenetic trees underlying this analysis are published in supplementary table 9 in Nagies *et al.* [20] and available under https://doi.org/10.25838/d5p-15. The distribution of verticality was generated for both the O_2-dependent and the O_2-independent reaction sets. The O_2-independent reaction set contained 3018 reactions and 8322 protein families with associated verticality values, while the O_2-dependent reaction set contained 365 reactions and 547 protein families with associated verticality values. Verticality values of all protein families associated with O_2-dependent and

O_2-independent reactions, as well as the average verticality of all protein families associated with each reaction (for O_2-dependent and O_2-independent reactions) were plotted against the number of genomes in which the protein family/reaction was detected (Fig. S1). Finally, we counted the number of O_2-dependent and O_2-independent reactions in the largest genome of each species in our dataset and applied regression models (see Fig. S2 and Table S3A).

Calculation of Gibbs energy and reactant count

The change in Gibbs energy ΔG was calculated for each reaction of both O_2-dependent and O_2-independent sets using eQuilibrator API [24] version 0.4.1. Calculations were performed for physiological conditions (pH 7, 1 mM concentrations of reactants and products, 25 °C, ionic strength 250 mM). In each reaction, O_2 was always written as a reactant (C00007 on the left side of the reaction), such that reaction R00009, R02550, R05229, and R00275 were reversed prior to calculation. In total, eQuilibrator yielded ΔG for 288 O_2-dependent reactions and for 2139 O_2-independent reactions (Table S4).

Reactants and products of O_2-dependent reactions were counted, with reactions written in the direction of O_2 consumption. For this count, only two ROS scavenging reactions were written in the O_2-evolving direction based on their physiological function: R00009 and R02670 (annotated as catalase). See Table S5 for the most common substrates and products of O_2-dependent reactions, and Table S6 for an overview of H_2O_2-consuming and H_2O_2-evolving reactions.

Identification of cofactors

Cofactors for each reaction were identified by integrating data from the IUBMB Comments section of the BRENDA database [25], the EC subclass descriptions if applicable, and literature data associated with each KO entry. In some cases, original literature not listed in KEGG was consulted, the references for which were included in Table S7. The BRENDA database was queried *via* EC number, while the KEGG literature data for each enzyme was accessed *via* the KOs associated with the corresponding reaction. In case of discrepancies, literature data was prioritized.

Only cofactors bound by the protein subunits corresponding to the KO annotation were listed. When there were multiple possibilities for the cofactors of an enzyme/subunit/chain, all were listed. Cytochromes and ferredoxins were listed as cofactors only if they were bound by the enzyme as soluble electron carriers. For cytochrome or ferredoxin enzyme domains, the cofactors listed were heme and iron–sulfur clusters, respectively. The reactions sometimes explicitly include an electron donor that provides electrons to the main enzyme indirectly, e.g., through an additional reductase component. Such electron donors were not listed, since they are not immediate ligands of the enzyme in question. In some cases, the natural electron donor was unknown and was therefore not listed. In addition, in cases where the cofactors were substrates or products in the reaction, they were not listed. Cosubstrates, such as 2-oxoglutarate in 2-oxoglutarate non-heme iron-dependent oxygenases, were also not listed.

Occurrence of functional categories per prokaryotic phylum

Using the KO identifiers, each protein family was assigned to at least one functional category (B-level) according to the KEGG BRITE classification. The distribution and occurrence of functional categories was examined for each phylum within all analyzed protein families linked to oxygen-dependent reactions (plotted in Fig. S3).

Statistical analysis

The calculated ΔG of the O_2-dependent and O_2-independent sets were compared with a *t*-test [26]. Comparisons of the number of O_2-dependent reactions per genome in aerobes and anaerobes and of genome size between aerobes and anaerobes (classified based on [27]) were done with Welch's *t*-test [26] (see Table S3B). In Fig. S2 the size of the largest genome of each species in the dataset was plotted against the number of O_2-dependent and O_2-independent reactions associated with the species. Regression models were chosen based on Pearson's product–moment correlation resulting in a linear model for O_2-dependent values and a logarithmic model for O_2-independent values (see Table S3A).

To compare the distributions of verticality values of O_2-dependent protein families and O_2-independent protein families, a Welch's *t*-test was performed on an unfiltered (Table S3C) and a filtered (Table S3D) dataset, in which high verticality values (> 1) were removed. This comparison was corroborated by subsampling the protein families associated with O_2-independent reactions. For this subsampling procedure, 100 000 samples of 547 O_2-independent protein families were generated to compare to the 547 protein families of O_2-dependent reactions. The average verticality (V_{avg}) was calculated for each sample. With this, an estimate was possible of how likely it would be to create a sample with a mean verticality as low as that of the O_2-dependent distribution from values in the O_2-independent distribution. This subsampling procedure was repeated, but changed each time to mitigate potential effects of unequal protein family size distributions. For this, four bins of equal size for protein family sizes (number of genomes in a protein family) of the O_2-dependent reactions were defined in the range of 4 genomes (minimum protein family size to

calculate phylogenetic trees) to 3380 genomes. For each bin, the number of protein families in the O_2-dependent distribution was counted, and during sampling, bins were filled up with an equal number of protein families from the O_2-independent distribution (see Table S3E).

Results

O_2 arose in a world without respiratory oxygen reductases

O_2-dependent enzymes are most frequent in facultative anaerobes with large genomes [28,29] (Tables S3, S8, and S9, Fig. S4). Plots of the frequency of O_2-dependent and O_2-independent reactions in relation to genome size are presented in Figs S2 and S5. Although respiratory oxygen reductases are not the focus of our study, it is important to keep their evolutionary significance in perspective. In today's environment, the overwhelming majority of O_2 produced by cyanobacteria and plants is used for respiration and energy conservation in aquatic and terrestrial environments, keeping the global carbon cycle in balance and our atmospheric O_2 levels largely constant [30], forest fires and human intervention notwithstanding. But that was clearly not the case before the GOE. The goal of our study is to gain insights into the role of O_2 for prokaryotes at the time of the GOE.

For the purpose of this study, we assume that at the time of the first appearance of environmental O_2 during the GOE, functioning respiratory oxygen reductases (or other O_2-dependent enzymes) had not yet evolved. Contrary to that view, it has been proposed that O_2 reductases evolved from a more ancient family of NO reductases, such that oxygen-utilizing terminal oxidases were already present when O_2 first appeared [31]. However, that proposal is not supported by current evidence [32-38], nor have similar proposals been put forward for other O_2-utilizing enzymes. There are three large and evolutionarily unrelated (independently arisen) superfamilies of oxygen reductases in respiratory chains: (a) the heme copper oxidases (HCO) [32,33] that include cytochrome *c* oxidases, (b) the heme-containing cytochrome *bd* oxidases that oxidize membrane quinols [34-36], and (c) the alternative oxidase superfamily (AOX) of non-heme diiron proteins that oxidize quinols [37,38]. Because members of the *bd* oxidase and AOX superfamilies only have one known electron acceptor substrate, O_2 can be directly inferred as the original substrate for the founding members of those enzyme families following the GOE. Only the HCO superfamily contains members that react with another electron acceptor substrate, the nitrous oxide (NO) reductases [27,33]. Recent surveys with broad sampling have clearly shown that NO reductases evolved multiple times and independently to produce the current spectrum of HCO terminal oxidases [33]. This indicates that O_2 was also the ancestral substrate for the HCO family, as in the case of the *bd* oxidases and AOX, in line with the low bioavailability of copper (an essential cofactor of HCO enzymes) prior to the origin of environmental O_2 [32].

This indicates, in turn, that terminal oxidases, like all other O_2-dependent enzymes, arose in environments that were experiencing a gradual encroachment of O_2 into an anoxic world. As O_2 first diffused into the environment, it led to oxidation of one-electron donors and enzyme inhibition, but it also introduced a novel, energy-rich oxidant into the trajectory of biochemical evolution. Once the HCO, *bd*, and AOX families of terminal oxidases arose, they rapidly diversified into new subfamilies [33,35,38]. Moreover, prokaryotes that possessed terminal oxidases came to flourish in heterotrophic settings, amplifying both the gene copy number and protein abundance of terminal oxidases in the environment over evolutionary time. Today terminal oxidases are, in terms of substrate turnover, unchallenged as the main consumers of O_2 in contemporary environments. In modern oceans, for example, photosynthetic oxygen production and respiratory O_2 consumption occur at roughly equal rates [39]. In mammals, 90% of oxygen consumption is mitochondrial respiration [40]. Respiration requires quantitatively large amounts of terminal oxidases, which are often complex membrane proteins.

In the present study our concern is not the increase of protein abundance from the GOE to today, rather, the situation at the onset of O_2 accumulation in Earth history and how O_2 could have impacted enzyme origin in the microbes that first encountered O_2. A situation very similar to that found in terminal oxidases is encountered with another O_2-utilizing (but not strictly O_2-dependent) enzyme, one involved in CO_2 fixation, RuBisCO. Today RuBisCO is not only the most common CO_2 fixing enzyme, it is the most abundant protein in nature [41]. The Calvin cycle in which it operates is the most recently evolved [42], the least energy efficient [43] and by far the most widespread of CO_2-fixing pathways. But it did not start that way. RuBisCO's original function was not in the Calvin cycle but in fermentative breakdown of RNA [44-46]. Changes in Earth's environment, in particular O_2, impacted the evolutionary trajectory of enzymes that play the central role in Earth's modern carbon cycle: terminal oxidases that consume O_2 and RuBisCO that fixes CO_2 with electrons from photosynthetic O_2

production. But what was the impact of O_2 on other enzymes and pathways in the immediate wake of the GOE? We asked O_2-dependent enzymes.

O_2 is an energy-rich compound, but unused in SLP

Oxygen is a strong oxidant, with a midpoint potential of +815 mV for the O_2/H_2O pair at pH 7. The enthalpy change in combustion reactions of O_2 with organic compounds (generating CO_2) is typically on the order of $-400\ kJ{\cdot}mol^{-1}$ of O_2, regardless of the organic compound undergoing combustion, because the energy released is almost entirely the energy stored in the energy-rich O_2 molecule, not in the organic reactant [7]. We estimated the free energy change, ΔG, for 288 of the 365 reactions that use O_2 as a substrate and that map to protein families (Table S1). The ΔG values are presented in Table S4 (along with the values estimated for 2139 O_2-independent reactions; see also Fig. S6), with an average calculated change in Gibbs free energy under physiological conditions of $-234\ kJ{\cdot}mol^{-1}$ of O_2. This value is sufficiently exergonic to drive substrate-level phosphorylation (SLP). It is therefore a curious observation that strictly O_2-dependent SLP reactions are virtually unknown. A change in free energy of roughly $-70\ kJ{\cdot}mol^{-1}$ is needed for the synthesis of one ATP [47], hence there is enough energy to synthesize at least one ATP (or more) per O_2-dependent oxidation of organic substrate on average. Out of 365 prokaryotic O_2-dependent reactions, only one (0.3% of the total) generates a product capable of supporting SLP: that catalyzed by the H_2O_2-producing (phosphorylating) pyruvate oxidase (EC 1.2.3.3) [48], POX. This enzyme converts pyruvate and O_2 to CO_2, H_2O_2 and acetyl phosphate with a calculated ΔG of $-158\ kJ{\cdot}mol^{-1}$. H_2O_2-producing pyruvate oxidase has a very narrow phylogenetic distribution, occurring almost exclusively among members of the *Lactobacilli*, and the role of O_2 is to oxidize a flavin in the reaction mechanism, not pyruvate itself. If O_2 were an evolutionary vehicle to improve energy yield—a widely held premise about the role of oxygen in evolution [49,50]—why would cells not conserve the ca. $-234\ kJ{\cdot}mol^{-1}$ of free energy released in > 99% of non-respiratory O_2-reducing reactions? The absence of O_2-dependent SLP reactions suggests that at the onset of the GOE, when O_2 first became available as a substrate, the main initial role of O_2 in microbial evolution was not immediately energy conservation, otherwise cells would likely have found ways to conserve the energy released in reactions of O_2 with organic compounds. The lack of O_2-dependent SLP reactions is noteworthy. If energy was not the initial functional role for oxygen, what then?

Oxygen-dependent enzymes often act on aromatic substrates

Gene-based studies addressing the physiological role of O_2 in evolution typically focus on its use as a terminal electron acceptor in respiratory chains [51,52], but terminal oxidases represent only about 1% of known prokaryotic O_2-utilizing enzyme families [28], hence they depict only one aspect of O_2 impact on prokaryotic metabolism. The responses of anaerobic microbes to O_2 *via* O_2 detoxification enzymes such as NADH oxidases and rubredoxin:oxygen oxidoreductase are well studied [6], as is the impact of O_2 on prokaryotic gene expression *via* DNA-binding proteins [53] such as the FeS cluster-containing O_2 sensor FNR (for fumarate and nitrate respiration) [54] and the ArcAB two-component system (for anoxic redox control or aerobic respiratory control), which responds to the oxidation state of the quinone pool [55]. We found that the most common substrates of O_2-dependent reactions are redox cofactors (ferredoxins, NAD(P)H, and flavins), water and 2-oxoglutarate, while common products include water, H_2O_2, CO_2, and ammonia (Tables S5 and S6).

The main utility of O_2 in biochemical reactions is that of a strong oxidant, affording microbes that possess the corresponding genes metabolic access to chemically stable substrates in the presence of O_2. Reactions of triplet O_2 in biological systems always involve a radical mechanism and almost always involve extraction of one electron from a substrate, generating a radical to initiate the mechanism [56]. The most frequent reaction types of prokaryotic O_2-dependent enzymes in this sample are aromatic degradation and amine oxidation (Table 1). Single electron donor dioxygenases, which incorporate both atoms of O_2 into the reaction product (EC 1.13.11.-), represent the most common enzyme category ($n = 57$), followed by NAD(P)H-dependent monooxygenases ($n = 56$), which incorporate one atom from O_2 into the product (EC 1.14.13.-). The enzymes from both these groups act mainly on aromatic substrates (Table 1). Dioxygenases typically disrupt aromatic rings [57]. The next most common enzymes are amine oxidases acting on CH–NH_2 groups (EC 1.4.3.-) ($n = 48$), copper or flavin containing proteins that catalyze the oxidation of primary amines, polyamines and amino acids [58]. NAD(P)H-dependent dioxygenases (EC 1.14.12.-), all of which act on aromatic substrates [59], are the next most common category ($n = 38$), followed by

Table 1. Most common prokaryotic O_2-dependent reactions. *n*: number of reactions; Dearo: number of dearomatizing reactions (the reaction destroys carbon ring aromaticity); Substr. Arom.: number of reactions where the substrate is an aromatic compound (excludes cosubstrates such as NAD(P)H, flavin, etc.); NH_3 Prod.: number of reactions releasing ammonia/ammonium ion as a reaction product; Dioxygenase: both oxygen atoms incorporated into substrate; Monooxygenase: one oxygen atom incorporated into substrate; Diverse: O_2 can be incorporated or reduced; Acceptor: O_2 is reduced, not incorporated.

EC number	*n*	e^- donor (substrate)	Fate of O_2 atoms	Dearo.	NH_3 prod.	Substr. Arom.
1.13.11.-	57	Single donor	Dioxygenase	40	0	45
1.14.13.-	56	Donor + NAD(P)H	Monooxygenase	4	0	38
1.4.3.-	48	CH-NH_2 group	Acceptor	0	27	21
1.14.12.-	38	Donor + NAD(P)H	Dioxygenase	28	2	38
1.14.15.-	28	Donor + FeS cluster	Monooxygenase	0	0	3
1.14.99.-	23	Misc. paired donors	Diverse	0	0	7
1.14.19.-	21	Paired donors	Acceptor	0	0	7
1.14.14.-	19	Flavin and Donor	Monooxygenase	0	0	6
1.1.3.-	14	CH-OH group	Acceptor	0	0	3
1.14.11.-	12	Paired donors	Diverse	0	0	2
1.5.3.-	8	CH-NH group	Acceptor	0	0	4
1.14.18.-	7	Paired donors	Monooxygenase	2	0	6
1.13.12.-	7	Single donor	Monooxygenase	0	0	4
1.14.20.-	6	Donor +2-OG	Diverse	0	0	2
1.10.3.-	6	Diphenols	Acceptor	2	0	5
1.3.3.-	5	CH-CH group	Diverse	2	0	3
7.1.1.-[a]	3	Quinones (or cyt *c*)	Acceptor	0	0	3

[a]Given out of order in the ranking, there are EC categories more common than the translocases (7.1.1.-, terminal oxidases) in the data, but the terminal oxidases are important, hence added to the list. EC numbers and links to all KEGG reactions are provided in Table S10.

FeS-dependent monooxygenases [60] ($n = 28$) (EC 1.14.15.-) (Table 1).

Of the 365 O_2-dependent KEGG reactions that mapped to protein families, more than 50% act on aromatics, stable substrates that require either a strong oxidant or a strong reductant [61] to disrupt the aromatic ring (Table 1). However, aromatic degradation in prokaryotes does not require O_2 [59,61,62]. Two O_2-independent routes of aromatic degradation *via* a benzoyl-CoA intermediate are catalyzed by unrelated benzoyl-CoA reductases. One route employs flavin-based electron bifurcation to generate midpoint potentials sufficiently negative to reduce benzoyl-CoA (E'_0 = –622 mV) [61]. Importantly, the growth rates of aerobic and anaerobic aromatic-degrading microbes are very similar, both having doubling times on the order of 4–6 h [59]. This indicates that the advantage conferred by the O_2-dependent pathway of aromatic degradation is not more rapid growth.

The evolutionary rationale behind the origin of O_2-dependent aromatic degradation might reside in the O_2-sensitivity of the evolutionarily older anaerobic enzymes. The O_2 sensitivity of solvent-exposed FeS clusters is a central underlying theme of O_2 in biochemical evolution [9,63–71]. The anaerobic ATP-dependent benzoyl-CoA reductase from *Thauera aromatica* contains three O_2-sensitive FeS clusters [72], while the electron-bifurcating benzoyl-CoA reductase enzyme complex from *Geobacter metallireducens* contains over 50 FeS clusters [61]. The more ancient, O_2-sensitive enzymes cannot function in oxic environments, requiring the origin of enzymes with an alternative reaction mechanism that can operate in oxic habitats [73]. The solvent-exposed nature of FeS clusters is important for their inhibition by O_2. Yet many FeS clusters in proteins are not solvent-exposed and not O_2-sensitive, such that the mere presence of an FeS cluster in a protein is not always a proxy for O_2 sensitivity. Human mitochondrial complex I contains eight FeS clusters, and photosystem I contains three, but neither is inhibited by O_2. In order for O_2 to oxidize an FeS cluster, it has to attain physical proximity, which is possible in the case of solvent-exposed clusters.

Oxygen-dependent enzymes confer novel physiological traits

The frequency of each enzyme (protein family) among the genomes in our sample is shown in Table 2 for the 30 most common O_2-dependent enzymes (the full list is given in Table S2). The most widespread O_2-dependent enzymes are either terminal oxidases, or enzymes employed in (cofactor) biosynthesis, detoxification, and substrate mobilization. In primary metabolism, the O_2-dependent biosynthetic reactions present

Table 2. Most widely distributed prokaryotic O_2-dependent enzymes. KO: identifier of the KEGG orthology group used for protein family annotation; N_g: number of genomes; *V*: verticality value (see Materials and methods). Enzyme function abbreviations are in brackets. Numbers in square brackets correspond to protons pumped (data from references [33,34]). The 7,8-dihydroneopterin oxygenase activity (K01633) is an oxygenase side reaction that proceeds through a carbanion intermediate, generating 7,8-dihydroxanthopterin, which is not a central intermediate in the folate synthesis pathway [74].

Enzyme name	EC number	KO	N_g	*V*
7,8-Dihydroneopterin oxygenase	1.13.11.81	K01633	4424	0.90
Cytochrome *bd* ubiquinol oxidase subunit II [0] (Oxphos)	7.1.1.7	K00425	4075	0.63
Cytochrome *bd* ubiquinol oxidase subunit I [0] (Oxphos)	7.1.1.7	K00426	4074	0.57
Catalase (O_2 detoxification)	1.11.1.6	K03781	3485	0.30
L-Aspartate oxidase (NAD+ synthesis; N mobilization)	1.4.3.16	K00278	3388	6.22
Pyridoxamine 5′-phosphate oxidase (PLP synthesis)	1.4.3.5	K00275	3026	0.11
Nitronate monooxygenase (N mobilization)	1.13.12.16	K00459	2974	0.34
Bacterioferritin (Iron storage)	1.16.3.1	K03594	2867	0.52
Glycolate dehydrogenase FAD-linked SU (2-OH acids)	1.1.99.14	K00104	2612	0.36
Cytochrome *o* ubiquinol oxidase subunit I [2] (Oxphos)	7.1.1.3	K02298	2561	0.14
Coproporphyrinogen III oxidase (Heme synthesis)	1.3.3.3	K00228	2473	1.18
Cytochrome *c* oxidase subunit I [4] (Oxphos)	7.1.1.9	K02274	2448	0.86
Cytochrome *c* oxidase subunit II [4] (Oxphos)	7.1.1.9	K02275	2435	0.39
2-polyprenylphenol 6-hydroxylase (UQ synthesis)	1.14.13.240	K18800	2431	1.49
Cytochrome *o* ubiquinol oxidase subunit II [2] (Oxphos)	7.1.1.3	K02297	2406	0.57
4,5-DOPA dioxygenase (Amino acids & aromatics ox.)	1.13.11.-	K15777	2266	1.15
Catalase-peroxidase (substrate oxidation and O_2 detox)	1.11.1.21	K03782	2197	0.14
Superoxide dismutase, Cu-Zn family (O_2 detoxification)	1.15.1.1	K04565	2180	1.07
2-octaprenyl-6-methoxyphenol hydroxylase (UQ synth.)	1.14.13.-	K03185	2157	0.72
Malate dehydrogenase, quinone (TCA cycle)	1.1.5.4	K00116	2095	0.31
Glycine oxidase (Thiamine biosynthesis)	1.4.3.19	K03153	1886	0.97
(S)-2-hydroxy-acid oxidase (2-OH acids, glycolate DH)	1.1.3.15	K11473	1849	0.90
Alkanesulfonate monooxygenase (S mobilization)	1.14.14.5	K00299	1831	0.09
Gamma-glutamyl putrescine oxidase (Amino acid ox.)	1.4.3.-	K09471	1809	0.09
tRNA 5-MAM-2-thiouridine bifunctional protein (tRNA)	1.5.-.-	K15461	1789	0.18
4-Hydroxyphenylpyruvate dioxygenase (Amino acid ox.)	1.13.11.27	K00457	1699	0.38
Alkanesulfonate monooxygenase (S mobilization)	1.14.14.5	K04091	1660	0.02
3-Phenylpropanoate dioxygenase (Amino acid ox.)	1.14.12.19	K00529	1540	0.28
4-Hydroxyphenylacetate 3-monooxygenase (AA ox.)	1.14.14.9	K00484	1539	0.09
Protoporphyrinogen oxidase (Heme synthesis)	1.3.3.4	K00231	1488	0.62
Cytochrome *c* oxidase subunit III [4] (Oxphos)	7.1.1.9	K02276	1433	0.49

post-GOE alternatives to O_2-independent reactions that existed in cells before the GOE. The oxidation of amino acids, sulfonates, or 2-hydroxy acids also present alternatives to preexisting anaerobic pathways.

The same principle holds for the HCO, *bd* and AOX superfamilies of terminal oxidases [27,51,75]. The family of cytochrome *bd* oxidases became integrated into previously-existing anaerobic electron transfer chains that used terminal acceptors such as sulfite and metals, producing a branching from membrane quinols to oxygen that is very common in extant facultatively anaerobic prokaryotes [34,51]. The appearance of HCOs introduced a branching in the electron transfer chains that pivot on *c*-type cytochromes and often use metals as source of electrons [76]. Large-scale phylogenetic studies of the HCO superfamily indicate that the most ancient HCOs are the A-type HCOs [33], which, like *bd* oxidases and AOX, normally do not reduce NO, as outlined above. The cytochrome *bd* oxidases are found among many archaeal groups and might be an ancient lineage of oxygen reductases, but their evolutionary history is, like that of HCOs, complicated by lateral gene transfers [32–38,75,76]. Though the relative ages of terminal oxidase families are not clarified because the protein families arose independently, and because the history of all three families has been affected by LGT [32–34,38], sulfide provides hints. HCO terminal oxidases require copper, which has very low bioavailability in sulfidic environments [75], and are strongly inhibited by sulfide [77,78], whereas *bd* oxidases are sulfide tolerant [79,80] and do not require copper. This would speak in favor of a greater antiquity for the *bd* family. The AOX family is probably the youngest of the three

oxygen reductase superfamilies, based on its very low affinity for O_2 [51].

In the present sample of prokaryotes, cytochrome *bd* ubiquinol oxidases appear to be the most common O_2-reducing terminal oxidases (Table 2). These transmembrane enzymes usually oxidize quinols but do not translocate protons across the membrane [34] although they do generate proton motive force *via* scalar protons (the localization of proton-consuming and proton-generating reactions on opposite sides of the membrane).

Another novel class of enzymes that arose in response to O_2 are detoxification enzymes that scavenge ROS (reactive oxygen species) that result from reactions of O_2: catalases [81], catalase-peroxidases [82], superoxide dismutases [83], and other detoxification enzymes [9], some of which are not represented in KEGG. The main cofactors of the ROS scavenging enzymes are metals [9,81–83]. Their main function is detoxification of reactive oxygen species such as the superoxide radical, $O_2^{\cdot-}$, that forms from one-electron transfers during O_2-dependent metabolism, or hydrogen peroxide, H_2O_2. Many novel enzymatic functions that did not exist prior to the existence of O_2 are found in the biosynthesis and degradation of secondary metabolites [28,84,85].

Genes for O_2-dependent enzymes undergo LGT more frequently than others

Evidence for O_2 accumulation in Earth history is geochemical (Fig. 1). Within that temporal framework, genes for O_2-dependent reactions can be inherited vertically, in which case their advantage is realized only within lineages, or they can be inherited *via* lateral gene transfer, in which case their advantage can be transmitted to any lineage. Several gene-based studies of O_2 in evolution address the timing of O_2 appearance [81,83,84] by using molecular clocks to date the age of O_2-dependent enzymes. Dating the appearance of O_2 with phylogenies of O_2-metabolizing enzymes requires, however, that the genes have not been subject to lateral gene transfer during evolution. A recent molecular-clock-based study of oxygen in evolution embraced the vertical view and ascribed the origin of oxygen-dependent enzymes to a single hypothetical lineage called the last universal oxygen ancestor [81]. In the real world of microbial genomes, it is known that all genes in prokaryotic genomes can be transferred and that most, if not all, have been subject to LGT at some point in evolution [20,85–87]. O_2-dependent enzymes are no exception.

If an O_2-dependent enzyme confers the ability to survive in oxic habitats, and if oxic habitats became more widespread across Earth history (Fig. 1), what is useful for one microbe can be useful for another. This predicts that O_2-dependent enzymes should readily spread across lineages *via* LGT during evolution. To quantify the role of LGT in genes for O_2-dependent enzymes, we utilized values of verticality, V, which provide a measure for how often a gene has been transferred between lineages in evolution [20]. Verticality measures only LGT events across major taxonomic boundaries (phyla or domains), evolutionarily recent fine-scale transfers are ignored in the calculation of verticality such that it provides a robust measure of LGT frequency [20]. High values of verticality indicate low levels of LGT for members of a gene family, for example ribosomal proteins, while low values of verticality reflect a high frequency of LGT for a given prokaryotic gene during evolution.

For the 547 protein families of O_2-dependent enzymes in prokaryotes for which we could assign a value of V, the mean verticality or V_{avg} is 0.273 ± 0.626 (avg $\pm$ SD). For comparison, the 3018 O_2-independent reactions in the data map to 11 754 protein families, of which 8322 had an associated verticality value (Table S1). The O_2-independent enzymes have a mean verticality of $V_{avg} = 0.781 \pm 1.926$. Although the means overlap, the difference in the two distributions is highly significant at $P = 7.43 \cdot 10^{-47}$ (t-value $= -14.92$, DF $= 1406$; Fig. 2). The difference remains significant even when higher verticality values (> 1) are filtered out ($V \leq 1$, t-value $= -6.22$, DF $= 603$, $P = 9.02 \cdot 10^{-10}$; see Table S3C,D).

Figure 2 shows that genes for O_2-dependent enzymes underwent more LGT in their evolution (expressed in lower verticality values) than genes for O_2-independent enzymes. Even the more widely distributed protein families and reactions have comparable low verticality (Fig. S1). The excess of more vertical protein families among O_2-independent enzymes (Fig. 2, upper right) is the signal of vertically evolving enzymes, many of them involved in information processing, such as aminoacyl-tRNA synthetases (for the complete list see Table S1). Genes for ribosomal proteins, though among the most vertically evolving genes in prokaryotes [20], are not plotted in Fig. 2 because their products do not catalyze chemical reactions, hence they are not represented in KEGG.

Figure 3 underscores the small contribution of vertical evolution in prokaryotic genes that code for enzymes involved in O_2-dependent reactions (blue points) relative to genes that are currently in use for molecular systematics studies, including ribosomal and other information processing proteins (black points),

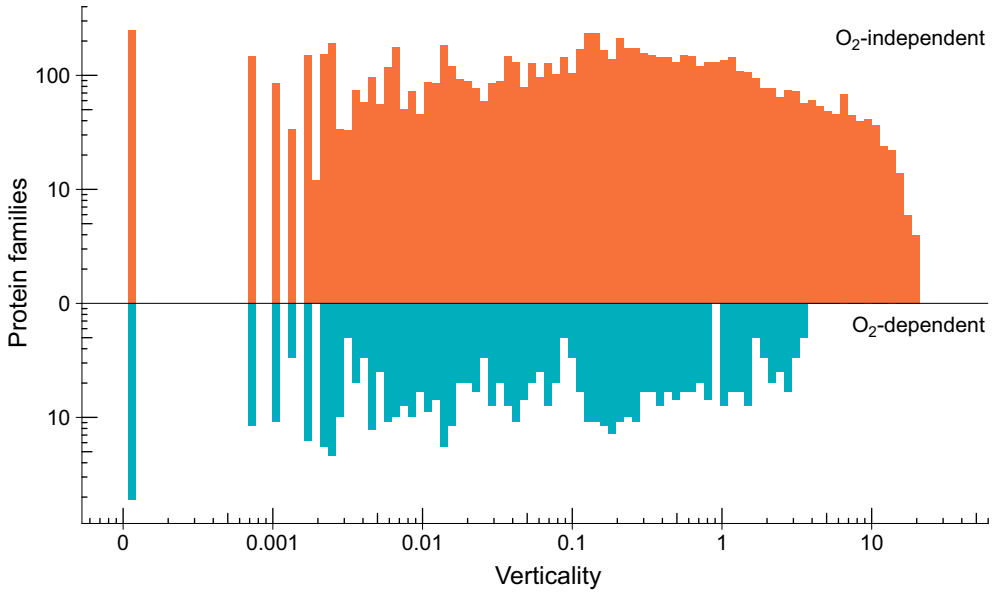


Fig. 2. Distribution of verticality values for protein families catalyzing oxygen-independent and oxygen-dependent reactions. The distribution of 8322 protein families with verticality values associated with O_2-independent reactions are shown in the upper histogram (orange) and the 547 protein families with verticality values associated with O_2-dependent reactions in the lower histogram (blue) in logarithmic scale, 100 bins. O_2-dependent gene families show a lower verticality, which means they have been subject to more LGT. For details of the statistical procedures see Materials and methods and Table S3.

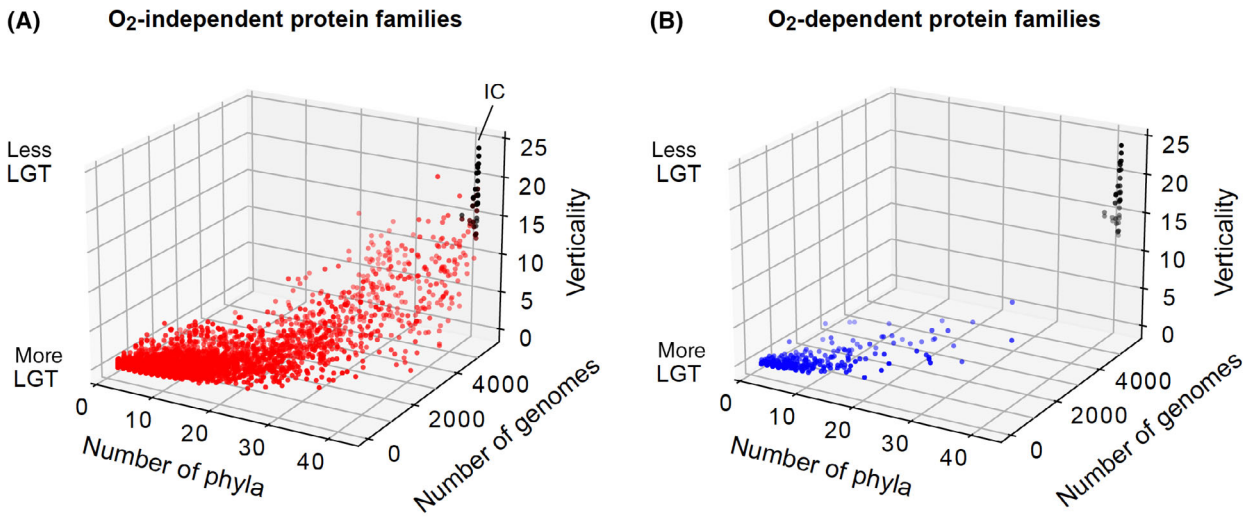
and genes that code for O_2-independent reactions (red points). The genes for O_2-dependent enzymes have almost no tendency at all to undergo vertical evolution, they are freely passed around lineages, where they can become fixed or not.

Within the 10 functional categories in KEGG with the highest frequency of O_2-dependent enzymes, the most frequently transferred O_2-dependent genes relative to O_2-independent genes within the same functional category encode products involved in the metabolism of 'other' amino acids (D-amino acids, glutathione, taurine, selenocompounds, etc.), followed by the functional category 'Protein families: metabolism' (which includes various catabolic reactions), genes coding for enzymes involved in xenobiotics degradation (includes cytochrome P_{450} enzymes), amino acid oxidation and breakdown of secondary metabolites (Fig. 3, Figs S3 and S7). In lipid metabolism, fatty acid and carotenoid oxidation, O_2-dependent enzymes are common, as these act on non-fermentable substrates. A summary of transfers across the nodes of a phylogenetic tree is presented in Fig. S8 (see also Table S11).

Evolutionary rationale behind O_2-dependent synthesis of essential cofactors

The most common genes for O_2-dependent enzymes we identified (Table 2) encode steps in the synthesis of essential cofactors including NAD^+, pyridoxal phosphate (PLP), heme, ubiquinone (UQ), and thiamine (Thi). There are also O_2-dependent and O_2-independent biosynthesis pathways for cobalamin and chlorophyll [73,88,89]. Why should organisms evolve O_2-dependent pathways for the biosynthesis of essential cofactors in the presence of preexisting O_2-independent pathways? One suggestion is that O_2-dependent pathways evolved in the presence of the older anaerobic pathways because they are favored for thermodynamic and kinetic reasons [90]. But the observation that microbial growth rates are similar for O_2-dependent and O_2-independent aromatic degradation [59] offers no hints that the thermodynamics or kinetics of the O_2-dependent pathways confer advantage. Why are there two kinds of pathways for seven essential cofactors?

In the case of parallel O_2-dependent and O_2-independent pathways for chlorophyll synthesis, Chew and Bryant [73] reached a conclusion that is probably applicable in most, if not all, cases of O_2-dependent and O_2-independent alternative pathways: "When a powerful selective pressure, oxygen, apparently inactivated oxygen-independent enzymes in (B)Chl biosynthesis, unrelated proteins with the same catalytic function but with completely different structures and mechanisms, sometimes even using oxygen as a substrate, then evolved." In other words, the inactivation



(C)

KEGG functional category	O_2-dependent Protein families	O_2-dependent V_{avg}	O_2-independent Protein families	O_2-independent V_{avg}	Ratio
Metabolism of other amino acids	27	0.065	548	0.622	9.55
Protein families: metabolism	41	0.124	1270	0.873	7.02
Xenobiotics biodegradation and metabolism	122	0.114	627	0.482	4.23
Amino acid metabolism	126	0.269	1804	0.818	3.04
Metabolism of other secondary metabolites	46	0.175	569	0.524	3.00
Lipid metabolism	41	0.155	1039	0.404	2.62
Energy metabolism	127	0.364	1138	0.911	2.50
Metabolism of cofactors and vitamins	75	0.552	1406	0.972	1.76
Carbohydrate metabolism	46	0.399	2564	0.589	1.48
Signal transduction	48	0.414	355	0.578	1.40

Fig. 3. Genes for enzymes catalyzing O_2-dependent reactions tend to be laterally transferred. Comparison of number of phyla in which a protein family is present, number of genomes where it occurs and verticality values for prokaryotic protein families. (A) The number of phyla (*x*-axis) is plotted in relation to the number of genomes (*z*-axis) and verticality values (*y*-axis) for 36 highly vertical and universal informational core genes (black) and for the 8322 O_2-independent protein families with associated verticality values (red). (B) The number of phyla (*x*-axis) is plotted in relation to the number of genomes (*z*-axis) and verticality values (*y*-axis) for 36 highly vertical and universal informational core genes (black) and the 547 protein families corresponding to oxygen-dependent reactions with associated verticality values (blue). The following functions correspond to the 36 informational core (IC) genes and are sorted by verticality: rpsJ, rpsK, rplA, alaS, rplB, ffh, rpsE, fusA, ftsY, hisS, truB, metG, rplN, pyrG, rpsH, rpsI, prsA, valS, rplE, rplF, tsaD, argS, truA, ychF, pyrH, uppS, ksgA, rpsS, serS, glyA, tuf, ileS, rpsL, eno, pgk, cyxS. (C) Average verticality values for O_2-dependent (blue points) and O_2-independent (red points) protein families across the 10 functional categories with the highest frequency of O_2-utilizing enzymes. The last column shows the ratio of average verticality between O_2-independent and O_2-dependent protein families per functional category. Values for all functional categories can be found in Table S12, and alternative plots showing the average verticality per functional category, including error bars, can be found in Fig. S7.

by O_2 of an anaerobe's preexisting enzyme generated the selection pressure for the evolution of an alternative enzyme that tolerated the presence of O_2, sometimes even by using O_2 in the radical-dependent mechanism. We suggest that the same evolutionary reason probably applies generally to the origin of O_2-dependent pathways for the synthesis of essential cofactors (or other essential functions) in the presence of preexisting O_2-independent pathways.

The O_2-independent pathways for essential cofactors must be older than the O_2-dependent pathways because prokaryotes that existed prior to the origin of oxygenic photosynthesis undoubtedly required and synthesized NAD^+ [91], PLP [92], ubiquinone, UQ [93,94], thiamine [95,96], heme [88], chlorophyll [73] and cobalamin [88,89]. The preexisting O_2-independent routes typically involve enzymes with O_2-sensitive low-potential metal centers, such as [4Fe4S] clusters,

O_2-sensitive pathway intermediates, or both. The reactions tend to rely on radical mechanisms, which are known to be common to many strict anaerobes [5]. Although the mechanism of O_2 activation is not known for many oxygen-dependent enzymes, the chemistry of O_2 in biological reactions prescribes that one-electron activation is necessary, which is why O_2 can interfere as an inhibitor of other biological radical reactions by extracting the radical. O_2-dependent biosynthetic pathways for essential cofactors offered microbes the tools needed to colonize oxic habitats, from which they would otherwise have been excluded. Enzymes with mechanisms of O_2 tolerance that do not involve O_2 as a substrate fall outside the scope of this paper, but the evolutionary rationale behind O_2 tolerance would also apply. The principle of oxygen's impact in evolution—inhibition first, then the ability to grow, then respiration—is illustrated in Fig. 4 (and highlighted in the title).

Flavins and iron are the most common cofactors across O_2-dependent reactions

As a consequence of its kinetically stable triplet diradical state, dioxygen has high activation energy barriers [4]. Because organic substrates usually exist in their singlet ground state, direct reactions of triplet oxygen with organic compounds are spin-forbidden [97] such that oxygen needs to be activated in order to react with typical organic substrates. In enzymatic reactions, activation is usually provided by the enzyme-guided donation of a single electron to generate a superoxide radical $O_2^{\cdot-}$ or by an electron and a proton to generate a perhydroxyl radical $HO_2^{\cdot}$. The one-electron donor is typically either a metal ion such as copper, iron (sometimes in the form of FeS centers), or manganese [98], or an organic cofactor such as a flavin or a pterin [99]. O_2 so activated by transfer of a single electron is extremely reactive, in O_2-dependent enzymes it readily oxidizes a specific substrate at the active site. The most common cofactors for the 365 reactions for which data could readily be identified are given in Table 3 (complete list in Table S7).

Physiological reactions of O_2 are radical reactions that require single electron donors to initiate the reaction with the O_2 diradical. This is reflected in the frequency of single electron donors (marker with an asterisk in Table 3) among the cofactors for O_2-dependent reactions sampled here. In some reactions, the substrate can provide the radical. In some reactions, the mechanism and the single electron source are unknown. Some of the cofactors are not directly involved in O_2 activation. The frequencies of the cofactors reflect their occurrence across individual reaction types, not their frequency within a protein or their occurrence in the environment. For example, a reaction catalyzed by a heme copper oxidase containing several hemes and several copper ions in all marine cyanobacteria (or in mammalian mitochondria, not sampled here) produces one count of heme and copper each. Similarly, the most abundant enzyme on Earth, RuBisCO, accounts for the lone occurrence of magnesium in the table.

The most common cofactors in O_2-dependent reactions are flavins and iron. Flavins are versatile coenzymes that perform both one- and two-electron transfers. They serve as cofactors and interact with dioxygen in several enzyme families, including flavin monooxygenases such as UbiH and UbiI involved in ubiquinone synthesis, PhzS from the phenazine pathway, or cyclohexanone monooxygenases ChnB. The most common activating mechanism of the flavin monooxygenases involves the reduced cofactor reacting with dioxygen to form the characteristic C(4a)-(hydro) peroxyflavin intermediate through a semiquinone radical pair [99,100].

Iron is the most common transition metal cofactor in oxygenases [101]. A common transition metal-dependent enzyme family in our set was the Rieske non-heme iron oxygenase family, including naphthalene-1,2-dioxygenase, the steroid hydroxylase KshAB involved in cholesterol catabolism, phthalate-4,5-dioxygenase and others. The oxygenase component of these enzymes is characterized by the presence of a catalytic mononuclear iron center and a Rieske iron–sulfur cluster. The latter accepts electrons from the two-electron donor NADH through a reductase component (often flavin-dependent), sometimes *via* an additional electron carrier. The electrons are transferred to the mononuclear iron center in order to activate dioxygen. The proposed mechanism includes rearrangements encompassing several oxidation states of the mononuclear iron center and a variety of radical and non-radical intermediates [102].

Flavins and iron are present individually, and sometimes together, in over 40% of the enzymes in our sample. NAD(P)H is very common as an electron donor (present in 23% of the reactions), providing electrons to oxygenases through flavins or the reductase component of a multi-component enzyme, as in the Rieske non-heme iron oxygenases. Iron–sulfur clusters are present as prosthetic groups in roughly 16% of the enzymes in our sample and about 17% bind the soluble one-electron carrier ferredoxin. Cytochromes as electron donors are rare for these

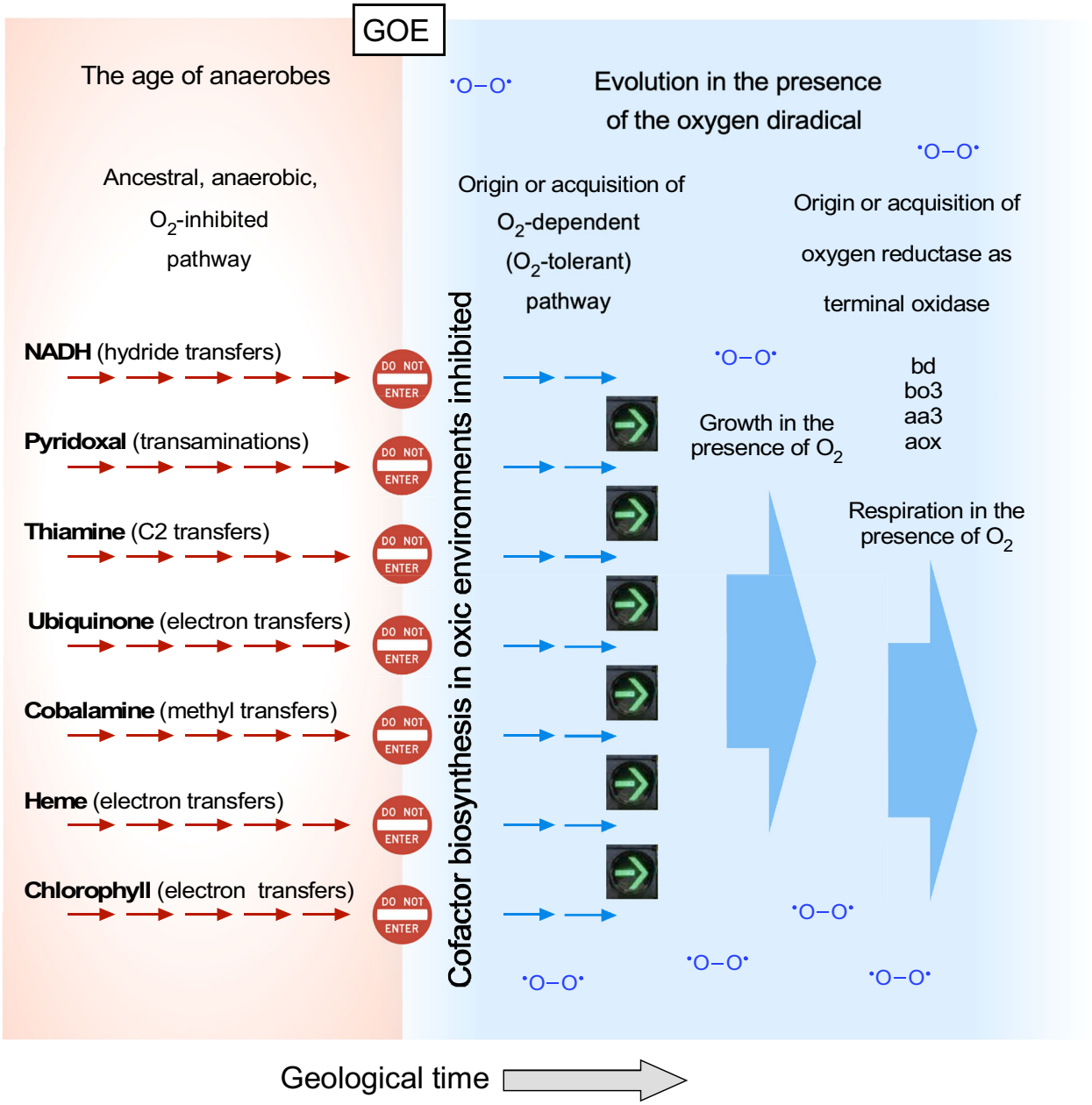


Fig. 4. Inhibition, growth, respiration. Several essential cofactors have O_2-inhibited and O_2-dependent biosynthetic pathways. The anaerobic pathways are older, obviously, because the cofactors shown are essential to various groups of prokaryotic anaerobes that existed prior to the origin of oxygenic photosynthesis. For references to the pathways see text. O_2-inhibited pathways preclude colonization of oxic niches ("do not enter" signs). Auxotrophy for the cofactor in an oxic niche is only an option if another group in the oxic niche has already evolved an O_2-tolerant biosynthetic route, in which case gene acquisition *via* LGT becomes an alternative to auxotrophy. The geological time of origin of oxygen respiration is not an issue here, only the relative timing of cofactor biosynthesis, which enables growth in an oxic environment, followed by origin or acquisition of O_2 respiratory terminal oxidases.

prokaryotic enzymes. The only O_2-dependent *S*-adenosyl methionine (SAM) binding enzyme identified in our sample was MnmC, a bifunctional enzyme found primarily in γ-Proteobacteria that catalyzes a specific modification of the tRNA wobble base by a mechanism including methylation preceded by oxidative cleavage [103].

Mononuclear iron centers and FeS clusters are often inhibited by oxygen [9]. This inhibition is likely the reason for the emergence of alternative pathways following

Table 3. The most common cofactors across 365 O_2-dependent reactions of prokaryotes. One-electron donors are marked with an (*).

Cofactors	Frequency
Flavins*	156
Iron*	152
NAD(P)+	84
Ferredoxin*	61
Iron–sulfur clusters*	57
Hemes*	39
Copper*	22
Quinones*	17
No cofactor or unknown	21
Coenzyme A	11
Cytochromes*	7
Pterins	5
Nickel	3
Flavodoxin*	3
Rubredoxin*	3
Manganese*	2
Tetrahydrofolate	2
Zinc	2
Selenium	1
Magnesium	1
S-adenosyl methionine*	1
Metal ion (unspecified)	1
Thiamine-pyrophosphate	1

the GOE. Why, then, are many alternative oxygen-tolerant enzymes replete with Fe and FeS clusters? The reactivity of oxygen with low-potential metal centers that results in cofactor inactivation in ancient O_2-independent enzymes, whose active sites are not adapted to dealing with oxygen, was not only circumvented, but rather used to good advantage for oxygen activation in the new, alternative pathways: once activated, O_2 can initiate a large number of radical reactions.

The KEGG list of O_2-dependent reactions in prokaryotes includes 21 having no associated cofactors, or where the cofactors were unknown. Several cofactor-independent oxygenases have recently been described [104–108]. Their mechanisms involve a substrate anion, generally generated by base catalysis, that donates a single electron to dioxygen, yielding a stabilized radical pair. The most familiar example of an oxygenase that requires no cofactors to activate dioxygen is the Calvin cycle enzyme RuBisCO, whose oxygenase mechanism involves a substrate-derived radical [107].

No evidence for pre-GOE environmental O_2 in genomes

Evidence for the appearance of O_2 in the atmosphere by the GOE 2.4 billion years ago is universally accepted and uncontroversial [1,2,10,11]. Some studies even date the GOE to a slightly more recent time, 2.3 billion years ago [108]. Several recent reports argue that atmospheric O_2 is older than the GOE. The proposals have it that geochemical interactions between quartz and water generated H_2O_2 early in Earth history and this quartz-derived H_2O_2 could have given rise to O_2 independent of the cyanobacterial OEC [109–111]. From a biological standpoint, an origin of O_2 from H_2O_2 seems extremely unlikely for several reasons. With a standard potential $E_o = +1.7$ V, H_2O_2 is a stronger oxidant than Cl_2 ($E_o = +1.36$ V) or O_2 ($E_o = +0.81$ V), hence a powerful disinfectant and general-purpose biocide, excluding a physical proximity between (so far unobserved) geological H_2O_2 production and life.

At the heart of H_2O_2-dependent abiotic O_2 origin models [109–111] is the chemical reaction catalyzed by catalase,

$$2\,H_2O_2 \rightarrow O_2 + 2\,H_2O$$

which is exothermic by $\Delta H = -101\ \mathrm{kJ{\cdot}mol^{-1}}$ of H_2O_2 at 100 °C [111] and exergonic with $\Delta G^{\circ\prime} = -189.9\ \mathrm{kJ{\cdot}mol^{-1}}$ (1 M H_2O_2, 25 °C, pH 7, 1 atm O_2) and $\Delta G_R = -154.2\ \mathrm{kJ{\cdot}mol^{-1}{\cdot}mol^{-1}}$ (for 1 mM H_2O_2, 100 °C, pH 7 and 0.1 atm O_2), making it an irreversible reaction under physiological conditions. The same reaction can be catalyzed by bifunctional catalase-peroxidases (Table S6), enzymes that degrade H_2O_2 either as a catalase (reaction above) or alternatively as a peroxidase [112] according to the reaction.

$$H_2O_2 + 2\,H - \mathrm{Substrate} \rightarrow 2\,H_2O + 2\,\mathrm{Substrate}$$

using the same active site, with ΔG depending on the substrate. Catalase is the most common H_2O_2 detoxifying enzyme in genomes (Table S6). Proponents of a sand/quartz origin of O_2 might argue that the abundance of catalases in genomes (Table S6) provides evidence in favor of an H_2O_2 origin of pre-GOE O_2, with catalase representing a widespread and ancient relict of a time when O_2 was made from H_2O_2. But catalase activity is essentially ubiquitous among prokaryotes, even among strict anaerobes [113], which generally lack O_2-dependent enzymes and pathways that could utilize O_2 (Figs S4 and S5). The function of catalase in strict anaerobes is to detoxify reactive oxygen species generated by one electron carriers during redox metabolism or by other organisms in the environment [77,114], not to provide anaerobes with O_2 from environmentally-supplied H_2O_2.

Furthermore, the proposed quartz-H_2O_2 mechanism of O_2 origin would, in principle, operate before the

origin of cells [111], in which case the genetic code, aminoacyl-tRNA synthetases, and ribosome biogenesis would have evolved in the presence of O_2 and H_2O_2. We found no O_2- or H_2O_2-dependent enzymes involved in core information processing functions. Thus, data from genomes and enzymes are distinctly at odds with suggestions for a pre-GOE (or prebiotic) origin of O_2 from H_2O_2 or quartz–water interactions.

Moreover, a recent incisive study by Koppenol and Sies [115] effectively rules out the possibility that abiotic H_2O_2 could have generated O_2 in amounts that would in any way impact the biosphere. They used simple and uncontroversial kinetics to calculate the expected half-life of H_2O_2 in the presence of 20–200 μM iron(II), a reasonable value for Archaean oceans. The key to their reasoning is the classical Fenton reaction, the rapid reaction of H_2O_2 with Fe^{2+} to yield Fe^{3+} and hydroxyl radicals, which rapidly react with almost any organic substrate. They obtain a conservative estimate for the half-life of H_2O_2 in the presence of Fe^{2+} as 0.7 s. Because Fe^{2+} was ubiquitous in all water-bearing environments on the early Earth, including the oceans, the miniscule 0.7 s half-life of H_2O_2 leaves no alternative to the conclusion that "before oxygenic photosynthesis, organisms were not exposed to H_2O_2, $HO^{\cdot}$, $O_2^{\cdot-}$, or $CO_3^{\cdot-}$" [115].

Another line of evidence suggesting pre-GOE O_2 comes from the possible presence of small amounts ("whiffs") of O_2 appearing in rocks 2.5 billion years of age, or ~ 50 MY prior to the GOE, based on the redox state of redox-sensitive minerals in sediments [116]. The evidence for pre-GOE whiffs of O_2 has recently been debated, however, with a new report indicating that the oxidation state of the molybdenum minerals originally supporting the existence of whiffs resulted from oxidation that took place after the GOE, long after the sediments had been deposited [117]. Those interpretations are challenged [118], and the challenge is in turn rebutted [119], suggesting anoxia (< 1 ppm O_2) in the sediment in question 2.5 billion years ago. From a biological perspective, the difference between oxygen appearance 2.4 billion years ago (the standard GOE model), or 50 million years earlier (the whiffs model), or 50 million years later [108] corresponds to a 2% difference in oxygen age, which has no impact on the verticality or functions (Fig. 3) of O_2-dependent enzymes surveyed here. It merely alters the time at which O_2 started to impact enzyme evolution by a factor of 0.02. In this study, we conservatively take the age of O_2 to correspond to the age of the GOE and we do not use molecular clocks.

Discussion

The main evolutionary impact of the origin of O_2 is traditionally viewed from the standpoint of eukaryote origin [49,84] or from the standpoint of mitochondrial respiration and animal evolution [50,120]. Yet eukaryotes appear roughly 1 billion years after the advent of O_2 [13,121], and animals appear in the fossil record near the base of the Cambrian, almost 2 billion years later than the appearance of O_2 [122,123]. Because both eukaryotes and animals emerged in a world that already contained biologically relevant amounts of O_2, their immediate ancestors were already equipped, enzymatically, to deal both with anoxia and with O_2 as a toxin and/or a substrate [124]. The fossilization of early invertebrates required the existence of O_2 for the synthesis of collagen by proline hydroxylases to generate hard body parts [125]. The origin of large animals also required the presence of O_2 [126]. Life on land was accompanied by adaptations to a permanently oxygenated atmosphere [127].

From the moment it was first produced by cyanobacteria, O_2 became an inhibitor of prokaryotic proteins with solvent-exposed FeS clusters, including ferredoxins, by acting as a one-electron acceptor in redox reactions [9,63]. The appearance of environmental O_2 following the GOE confronted prokaryotic life with a few challenges and numerous chemical opportunities. The main challenge was that a very small number of enzymes in anaerobes—sometimes only one enzyme per species, but often in physiologically key positions in metabolism—are inhibited by exposure to O_2 or activated forms thereof [6,9,63–73]. Such enzymes, for example pyruvate:ferredoxin oxidoreductase, a key enzyme in carbon and energy metabolism of anaerobes [65], or nitrogenase, the biosphere's N_2-fixing enzyme [128], typically have either a radical mechanism [5] or harbor low-potential metal centers, very often FeS centers, that can spontaneously react with the O_2 diradical [71], or both. The O_2-dependent inactivation of essential enzymes can arrest growth until anoxia is restored, which allows the organism to replace the poisoned enzyme by repair or resynthesis [6].

The evolutionary significance of O_2 is traditionally viewed in terms of energetic efficiency, O_2 having enabled improved ATP yield from heterotrophic substrate breakdown. The underlying reasoning is often that "life with oxygen is better than life without: a given amount of glucose processed in the presence of oxygen produces 18 times as much energy as the same amount of glucose processed without oxygen" [50]. *Escherichia coli* gains 15 ATP per glucose during O_2

respiration under optimal conditions and 4 ATP per glucose from anaerobic fermentation [129], the difference in maximal ATP yield is roughly a factor of 3.8, not 18.

Terrestrial, multicellular eukaryotes are, of course, highly specialized to an O_2-containing atmosphere and possess an O_2-dependent metabolism, yet this specialization took place relatively late in evolution [13,124,127]. Our present investigation deals exclusively with prokaryotes, as they were the only organisms to directly witness the advent of O_2 in evolution. In prokaryotes, O_2 presence does not always mean more energy, and more energy is not always in line with the physiological needs of the cell. The membrane potential, $\Delta\Psi$, used by *E. coli* for substrate import and ATP synthesis under anaerobic growth, −130 mV, is roughly the same as that under aerobic growth, −140 mV [130]. Yeast exhibits a similar response to *E. coli*'s overflow metabolism, called the Crabtree effect, the preference for fermentation in the presence of glucose and oxygen, which results in an increased rate of ATP production at low efficiency [18,131], not in an increased efficiency of ATP production.

In organisms that respire O_2, the diversity of terminal respiratory oxidases exhibits a variety of bioenergetic capacity. Among the enzymes of the HCO superfamily [51,75,76], the *aa3* type cytochrome *c* oxidases and the bo_3 type ubiquinol oxidases, both belonging to family A, pump 4 protons per O_2 reduced, while the cbb_3 type cytochrome *c* oxidases (family C) usually pump 2 protons per O_2 [35,75,132]. Among the growing number of oxidases belonging to the B family of HCO [33], there are subfamilies that pump 4 protons per O_2, various subfamilies that pump 2 protons per O_2 as the prototypic B-family oxidase of *Thermus* [132] and five subfamilies that do not pump protons, also because they do not react with oxygen for the lack of an active-site Tyr, which is involved in the formation of the His-Tyr cofactor in oxygen reductases. Conversely, the *bd* type ubiquinol oxidases and AOX pump 0 protons per O_2 [35,51], although the *bd* oxidases generate protonmotive force [34,35]. The subunits of non-pumping *bd* type oxidases are among the most widespread O_2-dependent enzymes in the present genome sample, apparently more widespread than those of the terminal oxidases that pump 2 or 4 protons per O_2 reduced (Table 2) (see references [33,34] for recent analyses of the HCO and *bd* oxidase superfamilies).

As with the Crabtree effect (2 ATP per glucose *via* cytosolic glycolysis in the presence of O_2), this spectrum of energy conservation within O_2-dependent terminal oxidases from 0 to 4 protons pumped per O_2 reduced suggests that the immediate physiological role of O_2 in evolution was not improved energy yield. Despite the highly exergonic nature of reactions between O_2 and organic compounds, the absence of proton-pumping capabilities in many terminal oxidases [34] and the noteworthy absence of O_2-dependent substrate-level phosphorylation suggest that the microbes that learned to harness O_2 in biochemical evolution were not limited in energy efficiency, but were instead limited in their ability to grow in Earth's increasingly oxic habitats because of enzymatic inhibition imposed by O_2.

Lateral gene transfer is the default mode of prokaryotic genome evolution [133]. It has impacted the distribution of O_2-dependent terminal oxidases across prokaryotic lineages [76,134]. Most genes in prokaryotic genomes have undergone LGT, with 97% of all genes having undergone at least one case of LGT between bacteria and archaea [135]. No gene family present in prokaryotic genomes has been completely immune to LGT between prokaryotic phyla during evolution [20]. We found that O_2-dependent enzymes underwent LGT more frequently than O_2-independent enzymes (Figs 2 and 3) and that this is true across almost all functional categories (Fig. 3 and Fig. S7, Table S12). The frequent spread of genes for O_2-dependent enzymes indicates that they conferred a physiological advantage to organisms that retained them. The nature of that advantage falls into three categories: (a) the breakdown of stable bonds in aromatic and nitrogenous compounds thereby mobilizing recalcitrant substrates in oxic environments, (b) the replacement (or supplementation) of enzymes that are inhibited by O_2, and (c) terminal oxidases (Table 2), but terminal oxidases can only be used by an organism once O_2-tolerant pathways of essential intermediate biosynthesis (or auxotrophy) have been incorporated into the genome (Fig. 4).

In primary metabolism, O_2-dependent enzymes have not generated fundamentally new biosynthetic or assimilatory traits. The central pathways of primary metabolism were in place and have remained largely unchanged since the time of the last universal common ancestor LUCA [136]. O_2-dependent enzymes presented alternative, O_2-tolerant routes to preexisting O_2-independent pathways that involved O_2-sensitive intermediates or O_2-sensitive cofactors, in particular solvent-exposed low-potential FeS clusters [64–71,73]. O_2-dependent enzymes are O_2-tolerant by nature, they allowed cells to colonize O_2-containing habitats from which they otherwise would have been excluded [73].

This factor underlies the higher rates of LGT we observe for O_2-dependent reactions. The colonization of oxic habitats not only required the presence of O_2 detoxification mechanisms [9,63], it required the presence of O_2-tolerant pathways for essential cofactor biosynthesis and substrate mobilization. O_2-tolerant enzymes that provide alternatives to the older O_2-sensitive counterparts in essential cofactor biosynthesis had to evolve and be expressible in the genome before the use of O_2 as a terminal acceptor in respiratory chains became an option for any microbe. Stated another way, O_2 is useless for a cell if the cell cannot grow in the presence of O_2 for lack of essential cofactors—NAD^+ [91], PLP [92], thiamine [95,96], quinones [93,94], chlorophyll [73], cobalamin [88,89], or heme [88]—whose anaerobic synthesis pathways are inhibited by O_2.

This selective pressure for the origin of O_2-tolerant and O_2-dependent enzymes, first suggested for chlorophyll biosynthesis [73], appears to apply very generally to O_2-dependent enzymes. Are there glaring exceptions to the rule? There is one, a big one—nitrogenase. Nitrogenase is rapidly inactivated by O_2 through damage to its solvent-exposed FeS clusters [128,137]. Yet in 4 billion years, life has never brought forth an O_2-dependent or O_2-tolerant alternative to Mo, Fe or V-dependent forms of nitrogenase. As a consequence, nitrogenase remained inhibited by atmospheric O_2 at ca. 1% PAL throughout the boring Pasteurian billion [12]. The inhibition of FeS clusters in nitrogenase by O_2 is well studied [128,137]. Nitrogenase inhibition by O_2 is a special case among O_2-dependent enzymes: it is the most important O_2-dependent enzyme that never evolved. Nitrogenase inhibition might have been the limiting factor for the increase in oxygen levels throughout the boring billion, through a negative feedback loop [12]. The essence of that negative feedback loop is that prior to the origin of cellulose synthesis by land plants, O_2 production by cyanobacteria and algae was stoichiometrically linked to synthesis of cell mass with a roughly 5 : 1 C : N ratio. Inhibition of nitrogenase brought biomass synthesis (hence O_2 production) to an immediate halt until environmental O_2 levels once again fell below 1% and nitrogenase could resume activity [12]. Even though nitrogenase inhibition by O_2 acted locally on FeS clusters in a protein, it exhibited an impact as global as cyanobacterial O_2 production itself. Our survey of O_2-dependent enzymes uncovers surprising generality of O_2 inhibition [73] and identifies it as a key mechanism that drove metabolic evolution, enabling prokaryotes to colonize oxic habitats, governing the composition of Earth's atmosphere [12] for almost 2 billion years.

Acknowledgements

We thank Karl Kleinermanns (Düsseldorf) and Yannick De Decker (Brussels) for their calculations of Gibbs free energy change for the catalase reaction. We thank Rebecca E. Gerhards for help with preparing the manuscript. We thank the Zentrum für Informations- und Medientechnologie (ZIM) of the HHU Düsseldorf for providing computational resources. This paper is dedicated to the memory of Dan Tawfik. This work was supported by the European Research Council (ERC) under the Horizon 2020 research and innovation program (grant 101018894 to WFM), the Volkswagen Foundation (grant 96742 to WFM), the Deutsche Forschungsgemeinschaft (grant MA 1426/21-1 to WFM) and the German-Israeli Project Cooperation (DIP) (grant 1426/23-1 to WFM and grant 2476/2-1 to IM). MM is supported by the Scientific Grant Agency of the Ministry of Education of the Slovak Republic (VEGA) (grant 1/0457/24). Open Access funding enabled and organized by Projekt DEAL.

Author contributions

WFM, NM, FSPN, JLEW and MM conceived the study. WFM supervised the study. WFM, NM, FSPN, JLEW, NK and MRK designed the research. NM, FSPN, JLEW, NK, MRK and KT curated the data. NM, FSPN, JLEW, NK, MRK, KT, LM and NB performed the bioinformatical analysis. WFM, NM, FSPN, JLEW, MRK, KT, LM and NB visualized the data. WFM, JFA, NM, FSPN, JLEW, NK, MRK, KT, LM, NB, MM, MDE and IM analyzed and interpreted the data. WFM, NM, FSPN, JLEW and MRK wrote the original manuscript. WFM, JFA, NM, FSPN, JLEW, NK, MRK, KT, LM, NB, MM, MDE and IM revised and edited the manuscript. WFM and IM acquired funding.

Peer review

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1002/1873-3468.14906.

Data accessibility

All data are available in the main text or the Supporting Information. The protein families used for this

analysis are available under https://www.molevol.hhu.de/resources.

References

1 Holland HD (2002) Volcanic gases, black smokers, and the great oxidation event. *Geochim Cosmochim Acta* **66**, 3811–3826.
2 Fischer WW, Hemp J and Johnson JE (2016) Evolution of oxygenic photosynthesis. *Annu Rev Earth Planet Sci* **44**, 647–683.
3 Demoulin CF, Lara YJ, Lambion A and Javaux EJ (2024) Oldest thylakoids in fossil cells directly evidence oxygenic photosynthesis. *Nature* **625**, 529–534.
4 Borden WT, Hoffmann R, Stuyver T and Chen B (2017) Dioxygen: what makes this triplet diradical kinetically persistent? *J Am Chem Soc* **139**, 9010–9018.
5 Buckel W and Golding BT (2006) Radical enzymes in anaerobes. *Annu Rev Microbiol* **60**, 27–49.
6 Lu Z and Imlay JA (2021) When anaerobes encounter oxygen: mechanisms of oxygen toxicity, tolerance and defense. *Nat Rev Microbiol* **19**, 774–785.
7 Schmidt-Rohr K (2015) Why combustions are always exothermic, yielding about 418 kJ per mole of O_2. *J Chem Educ* **92**, 2094–2099.
8 Brewer L (1952) The thermodynamic properties of the oxides and their vaporization processes. *Chem Rev* **52**, 1–75.
9 Khademian M and Imlay JA (2021) How microbes evolved to tolerate oxygen. *Trends Microbiol* **29**, 428–440.
10 Lyons TW, Reinhard CT and Planavsky NJ (2014) The rise of oxygen in Earth's early ocean and atmosphere. *Nature* **50**, 307–315.
11 Lenton TM, Dahl TW, Daines SJ, Mills BJW, Ozaki K, Saltzman MR and Porada P (2016) Earliest land plants created modern levels of atmospheric oxygen. *Proc Natl Acad Sci USA* **113**, 9704–9709.
12 Allen JF, Thake B and Martin WF (2019) Nitrogenase inhibition limited oxygenation of Earth's proterozoic atmosphere. *Trends Plant Sci* **24**, 1022–1031.
13 Mills DB, Boyle RA, Daines SJ, Sperling EA, Pisani D, Donoghue PCJ and Lenton TM (2022) Eukaryogenesis and oxygen in Earth history. *Nat Ecol Evol* **6**, 520–532.
14 Mukherjee I, Large RR, Corkrey R and Danyushevsky LV (2018) The boring billion, a slingshot for complex life on Earth. *Sci Rep* **8**, 4432.
15 Mojzsis SJ, Arrhenius G, McKeegan KD, Harrison TM, Nutman AP and Friend CRL (1996) Evidence for life on Earth before 3,800 million years ago. *Nature* **384**, 55–59.
16 Blaser MB, Dreisbach LK and Conrad R (2013) Carbon isotope fractionation of 11 acetogenic strains grown on H_2 and CO_2. *Appl Environ Microbiol* **79**, 1787–1794.
17 Arndt NT and Nisbet EG (2012) Processes on the young Earth and the habitats of early life. *Annu Rev Earth Planet Sci* **40**, 521–549.
18 Martin WF, Tielens AGM and Mentel M (2020) Mitochondria and Anaerobic Energy Metabolism in Eukaryotes: Biochemistry and Evolution. De Gruyter, Berlin.
19 Kanehisa M and Goto S (2000) KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**, 27–30.
20 Nagies FSP, Brueckner J, Tria FDK and Martin WF (2020) A spectrum of verticality across genes. *PLoS Genet* **16**, e1009200.
21 Enright AJ, van Dongen S and Ouzounis CA (2002) An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* **30**, 1575–1584.
22 Brueckner J and Martin WF (2020) Bacterial genes outnumber archaeal genes in eukaryotic genomes. *Genome Biol Evol* **12**, 282–292.
23 Buchfink B, Xie C and Huson D (2015) Fast and sensitive protein alignment using DIAMOND. *Nat Methods* **12**, 59–60.
24 Flamholz A, Noor E, Bar-Even A and Milo R (2012) eQuilibrator—the biochemical thermodynamics calculator. *Nucleic Acids Res* **40**, D770–D775.
25 Chang A, Jeske L, Ulbrich S, Hofmann J, Koblitz J, Schomburg I, Neumann-Schaal M, Jahn D and Schomburg D (2021) BRENDA, the ELIXIR core data resource in 2021: new developments and updates. *Nucleic Acids Res* **49**, D498–D508.
26 Welch BL (1947) The generalization of "Student's" problem when several different population variances are involved. *Biometrika* **34**, 28–35.
27 Sousa FL, Alves RJ, Pereira-Leal JB, Teixeira M and Pereira MM (2011) A bioinformatics classifier and database for heme-copper oxygen reductases. *PLoS One* **6**, e19117.
28 Sousa FL, Nelson-Sathi S and Martin WF (2016) One step beyond a ribosome: the ancient anaerobic core. *Biochim Biophys Acta* **1857**, 1027–1038.
29 Jabøońska J and Tawfik DS (2019) The number and type of oxygen-utilizing enzymes indicates aerobic vs. anaerobic phenotype. *Free Radic Biol Med* **140**, 84–92.
30 Li C, Huang J, Ding L, Liu X, Han D and Huang J (2021) Estimation of oceanic and land carbon sinks based on the most recent oxygen budget. *Earth's Future* **9**, e2021EF002124.
31 Ducluzeau A-L, van Lis R, Duval S, Schoepp-Cothenet B, Russell MJ and Nitschke W (2009) Was nitric oxide the first deep electron sink? *Trends Biochem Sci* **34**, 9–15.
32 Pereira MM, Santana M and Teixeira M (2001) A novel scenario for the evolution of haem–copper oxygen reductases. *Biochim Biophys Acta* **1505**, 185–208.

33 Murali R, Hemp J and Gennis RB (2022) Evolution of quinol oxidation within the heme-copper oxidoreductase superfamily. *Biochim Biophys Acta Bioenerg* **1863**, 148907.

34 Murali R, Gennis RB and Hemp J (2021) Evolution of the cytochrome bd oxygen reductase superfamily and the function of CydAA' in Archaea. *ISME J* **15**, 3534–3548.

35 Borisov VB, Gennis RB, Hemp J and Verkhovsky MI (2011) The cytochrome bd respiratory oxygen reductases. *Biochim Biophys Acta Bioenerg* **1807**, 1398–1413.

36 Degli Esposti M, Rosas-Pérez T, Servín-Garcidueñas LE, Bolaños LM, Rosenblueth M and Martínez-Romero E (2015) Molecular evolution of cytochrome bd oxidases across proteobacterial genomes. *Genome Biol Evol* **7**, 801–820.

37 Atteia A, van Lis R, van Hellemond JJ, Tielens AGM, Martin W and Henze K (2004) Identification of prokaryotic homologues indicates an endosymbiotic origin for the alternative oxidases of mitochondria (AOX) and chloroplasts (PTOX). *Gene* **330**, 143–148.

38 Pennisi R, Salvi D, Brandi V, Angelini R, Ascenzi P and Polticelli F (2016) Molecular evolution of alternative oxidase proteins: a phylogenetic and structure modeling approach. *J Mol Evol* **82**, 207–218.

39 del Giorgio PA and Duarte CM (2002) Respiration in the open ocean. *Nature* **420**, 379–384.

40 Rolfe DF and Brown GC (1997) Cellular energy utilization and molecular origin of standard metabolic rate in mammals. *Physiol Rev* **77**, 731–758.

41 Erb TJ and Zarzycki J (2018) A short history of RubisCO: the rise and fall (?) of Nature's predominant CO_2 fixing enzyme. *Curr Opin Biotechnol* **49**, 100–107.

42 Fuchs G (2011) Alternative pathways of carbon dioxide fixation: insights into the early evolution of life? *Annu Rev Microbiol* **65**, 631–658.

43 Berg IA, Kockelkorn D, Ramos-Vera WH, Say RF, Zarzycki J, Hügler M, Alber BE and Fuchs G (2010) Autotrophic carbon fixation in archaea. *Nat Rev Microbiol* **8**, 447–460.

44 Aono R, Sato T, Yano A, Yoshida S, Nishitani Y, Miki K, Imanaka T and Atomi H (2012) Enzymatic characterization of AMP phosphorylase and ribose-1,5-bisphosphate isomerase functioning in an archaeal AMP metabolic pathway. *J Bacteriol* **194**, 6847–6855.

45 Aono R, Sato T, Imanaka T and Atomi H (2015) A pentose bisphosphate pathway for nucleoside degradation in Archaea. *Nat Chem Biol* **11**, 355–360.

46 Schönheit P, Buckel W and Martin WF (2016) On the origin of heterotrophy. *Trends Microbiol* **24**, 12–25.

47 Thauer RK, Jungermann K and Decker K (1977) Energy conservation in chemotrophic anaerobic bacteria. *Bacteriol Rev* **41**, 100–180.

48 Muller YA and Schulz GE (1993) Structure of the thiamine- and flavin-dependent enzyme pyruvate oxidase. *Science* **259**, 965–967.

49 Margulis L (1970) Origin of Eukaryotic Cells. Yale University Press, New Haven, CT.

50 Rytkönen KT (2018) Evolution: oxygen and early animals. *Elife* **7**, e34756.

51 Degli Esposti M, Mentel M, Martin W and Sousa FL (2019) Oxygen reductases in alphaproteobacterial genomes: physiological evolution from low to high oxygen environments. *Front Microbiol* **10**, 499.

52 Brochier-Armanet C, Talla E and Gribaldo S (2009) The multiple evolutionary history of dioxygen reductases: implications for the origin and evolution of aerobic respiration. *Mol Biol Evol* **26**, 285–297.

53 Allen JF (1993) Redox control of transcription – sensors, response regulators, activators and repressors. *FEBS Lett* **332**, 203–207.

54 Unden G and Bongaerts J (1997) Alternative respiratory pathways of *Escherichia coli*: energetics and transcriptional regulation in response to electron acceptors. *Biochim Biophys Acta Bioenerg* **1320**, 217–234.

55 Brown AN, Anderson MT, Bachman MA and Mobley HLT (2022) The ArcAB two-component system: function in metabolism, redox control, and infection. *Microbiol Mol Biol Rev* **86**, e00110-21.

56 Fridovich I (1989) Superoxide dismutases. An adaptation to a paramagnetic gas. *J Biol Chem* **264**, 7761–7764.

57 Vaillancourt FH, Bolin JT and Eltis LD (2006) The ins and outs of ring-cleaving dioxygenases. *Crit Rev Biochem Mol Biol* **41**, 241–267.

58 Gaweska H and Fitzpatrick PF (2011) Structures and mechanism of the monoamine oxidase family. *Biomol Concepts* **2**, 365–377.

59 Fuchs G, Boll M and Heider J (2011) Microbial degradation of aromatic compounds – from one strategy to four. *Nat Rev Microbiol* **9**, 803–816.

60 Vanoni MA (2021) Iron-sulfur flavoenzymes: the added value of making the most ancient redox cofactors and the versatile flavins work together. *Open Biol* **11**, 210010.

61 Huwiler SG, Löffler C, Anselmann SEL, Stärk HJ, von Bergen M, Flechsler J, Rachel R and Boll M (2019) One-megadalton metalloenzyme complex in *Geobacter metallireducens* involved in benzene ring reduction beyond the biological redox window. *Proc Natl Acad Sci USA* **116**, 2259–2264.

62 Fuchs G (2008) Anaerobic metabolism of aromatic compounds. *Ann N Y Acad Sci* **1125**, 82–99.

63 Imlay JA (2003) Pathways of oxidative damage. *Annu Rev Microbiol* **57**, 395–418.

64 Schlesier J, Rohde M, Gerhardt S and Einsle O (2016) A conformational switch triggers nitrogenase

protection from oxygen damage by Shethna protein II (FeSII). *J Am Chem Soc* **138**, 239–247.
65 Ragsdale SW (2003) Pyruvate ferredoxin oxidoreductase and its radical intermediate. *Chem Rev* **103**, 2333–2346.
66 Boyd ES, Thomas KM, Dai Y, Boyd JM and Outten FW (2014) Interplay between oxygen and Fe-S cluster biogenesis: insights from the Suf pathway. *Biochemistry* **53**, 5834–5847.
67 Khoroshilova N, Popescu C, Münck E, Beinert H and Kiley PJ (1997) Iron-sulfur cluster disassembly in the FNR protein of *Escherichia coli* by O_2: [4Fe-4S] to [2Fe-2S] conversion with loss of biological activity. *Proc Natl Acad Sci USA* **94**, 6087–6092.
68 Pan N and Imlay JA (2001) How does oxygen inhibit central metabolism in the obligate anaerobe Bacteroides thetaiotaomicron. *Mol Microbiol* **39**, 1562–1571.
69 Orme-Johnson WH and Beinert H (1969) On the formation of the superoxide anion radical during the reaction of reduced iron-sulfur proteins with oxygen. *Biochem Biophys Res Commun* **36**, 905–911.
70 Allen JF (1975) A two-step mechanism for the photosynthetic reduction of oxygen by ferredoxin. *Biochem Biophys Res Commun* **66**, 36–43.
71 Imlay JA (2006) Iron-sulphur clusters and the problem with oxygen. *Mol Microbiol* **59**, 1073–1082.
72 Tiedt O, Fuchs J, Eisenreich W and Boll M (2018) A catalytically versatile benzoyl-CoA reductase, key enzyme in the degradation of methyl- and halobenzoates in denitrifying bacteria. *J Biol Chem* **293**, 10264–10274.
73 Chew AGM and Bryant DA (2007) Chlorophyll biosynthesis in bacteria: the origins of structural and functional diversity. *Annu Rev Microbiol* **61**, 113–129.
74 Czekster CM and Blanchard JS (2012) One substrate, five products: reactions catalyzed by the dihydroneopterin aldolase from *Mycobacterium tuberculosis*. *J Am Chem Soc* **134**, 19758–19771.
75 Sousa FL, Alves RJ, Ribeiro MA, Pereira-Leal JB, Teixeira M and Pereira MM (2012) The superfamily of heme-copper oxygen reductases: types and evolutionary considerations. *Biochim Biophys Acta* **1817**, 629–637.
76 Degli Esposti M (2020) On the evolution of cytochrome oxidases consuming oxygen. *Biochim Biophys Acta Bioenerg* **1861**, 148304.
77 Nicholls P (1975) The effect of sulfide on cytochrome aa3. Isosteric and allosteric shifts of the reduced α-peak. *Biochim Biophys Acta* **396**, 24–35.
78 Nicholls P, Marshall DC, Cooper CE and Wilson MT (2013) Sulfide inhibition of and metabolism by cytochrome c oxidase. *Biochem Soc Trans* **41**, 1312–1316.
79 Forte E, Borisov VB, Falabella M, Colaço HG, Tinajero-Trejo M, Poole RK, Vicente JB, Sarti P and Giuffrè A (2016) The terminal oxidase cytochrome bd promotes sulfide-resistant bacterial respiration and growth. *Sci Rep* **6**, 23788.
80 Korshunov S, Imlay KRC and Imlay JA (2016) The cytochrome *bd* oxidase of *Escherichia coli* prevents respiratory inhibition by endogenous and exogenous hydrogen sulfide. *Mol Microbiol* **101**, 62–77.
81 Tehrani HS and Moosavi-Movahedi AA (2018) Catalase and its mysteries. *Prog Biophys Mol Biol* **140**, 5–12.
82 Khmelevtsova LE, Sazykin IS, Azhogina TN and Sazykina MA (2020) Prokaryotic peroxidases and their application in biotechnology (review). *Appl Biochem Microbiol* **56**, 373–380.
83 Bafana A, Dutt S, Kumar A, Kumar S and Ahuja PS (2011) The basic and applied aspects of superoxide dismutase. *J Mol Catal B: Enzym* **68**, 129–138.
84 Raymond J and Segrè D (2006) The effect of oxygen on biochemical networks and the evolution of complex life. *Science* **311**, 1764–1767.
85 Jabøońska J and Tawfik DS (2021) The evolution of oxygen-utilizing enzymes suggests early biosphere oxygenation. *Nat Ecol Evol* **5**, 442–448.
86 Dagan T and Martin W (2007) Ancestral genome sizes specify the minimum rate of lateral gene transfer during prokaryote evolution. *Proc Natl Acad Sci USA* **104**, 870–875.
87 Dagan T, Artzy-Randrup Y and Martin W (2008) Modular networks and cumulative impact of lateral transfer in prokaryote genome evolution. *Proc Natl Acad Sci USA* **105**, 10039–10044.
88 Dailey HA, Dailey TA, Gerdes S, Jahn D, Jahn M, O'Brian MR and Warren MJ (2017) Prokaryotic heme biosynthesis: multiple pathways to a common essential product. *Microbiol Mol Biol Rev* **81**, e00048-16.
89 Bryant DA, Hunter CN and Warren MJ (2020) Biosynthesis of the modified tetrapyrroles—the pigments of life. *J Biol Chem* **295**, 6888–6925.
90 Raymond J and Blankenship RE (2005) Biosynthetic pathways, gene replacement and the antiquity of life. *Geobiology* **2**, 199–203.
91 Ollagnier-de Choudensa S, Loiseau L, Sanakis Y, Barras F and Fontecave M (2005) Quinolinate synthetase, an iron–sulfur enzyme in NAD biosynthesis. *FEBS Lett* **579**, 3737–3743.
92 Mukherjee T, Hanes J, Tews I, Ealick SE and Begley TP (2011) Pyridoxal phosphate: biosynthesis and catabolism. *Biochim Biophys Acta Proteins Proteomics* **1814**, 1585–1596.
93 Degli Esposti M (2017) A journey across genomes uncovers the origin of ubiquinone in cyanobacteria. *Genome Biol Evol* **9**, 3039–3053.
94 Pelosi L, Vo CD, Abby SS, Loiseau L, Rascalou B, Hajj Chehade M, Faivre B, Goussé M, Chenal C, Touati N *et al.* (2019) Ubiquinone biosynthesis over

the entire O_2 range: characterization of a conserved O_2-independent pathway. *MBio* **10**, e01319-19.
95 Leonardi R, Fairhurst SA, Kriek M, Lowe DJ and Roach PL (2003) Thiamine biosynthesis in *Escherichia coli*: isolation and initial characterisation of the ThiGH complex. *FEBS Lett* **539**, 95–99.
96 Settembre EC, Dorrestein PC, Park JH, Augustine AH, Begley TP and Ealick SE (2003) Structural and mechanistic studies on ThiO, a glycine oxidase essential for thiamin biosynthesis in *Bacillus subtilis*. *Biochemistry* **42**, 2971–2981.
97 Klinman JP (2001) Life as aerobes: are there simple rules for activation of dioxygen by enzymes? *J Biol Inorg Chem* **6**, 1–13.
98 Huang X and Groves JT (2018) Oxygen activation and radical transformations in heme proteins and metalloporphyrins. *Chem Rev* **118**, 2491–2553.
99 Romero E, Gómez Castellanos JR, Gadda G, Fraaije MW and Mattevi A (2018) Same substrate, many reactions: oxygen activation in flavoenzymes. *Chem Rev* **118**, 1742–1769.
100 Wongnate T, Surawatanawong P, Visitsatthawong S, Sucharitakul J, Scrutton NS and Chaiven P (2014) Proton-coupled electron transfer and adduct configuration are important for C4a-hydroperoxyflavin formation and stabilization in a flavoenzyme. *J Am Chem Soc* **136**, 241–253.
101 Wang Y, Li J and Liu A (2017) Oxygen activation by mononuclear nonheme iron dioxygenases involved in the degradation of aromatics. *J Biol Inorg Chem* **22**, 395–405.
102 Barry SM and Challis GL (2013) Mechanism and catalytic diversity of Rieske non-heme iron-dependent oxygenases. *ACS Catal* **3**, 2362–2370.
103 Kim J and Almo SC (2013) Structural basis for hypermodification of the wobble uridine in tRNA by bifunctional enzyme MnmC. *BMC Struct Biol* **13**, 1–13.
104 Widboom PF, Fielding EN, Liu Y and Bruner SD (2007) Structural basis for cofactor-independent dioxygenation in vancomycin biosynthesis. *Nature* **447**, 342–345.
105 Frerichs-Deeken U, Ranguelova K, Kappl R, Hüttermann J and Fetzner S (2004) Dioxygenases without requirement for cofactors and their chemical model reaction: compulsory order ternary complex mechanism of 1 H-3-hydroxy-4-oxoquinaldine 2, 4-dioxygenase involving general base catalysis by histidine 251 and single-electron oxidation of the substrate dianion. *Biochemistry* **43**, 14485–14499.
106 Baas BJ, Poddar H, Geertsema EM, Rozeboom HJ, de Vries MP, Permentier HP, Thunnissen AMWH and Poelarends GJ (2015) Functional and structural characterization of an unusual cofactor-independent oxygenase. *Biochemistry* **54**, 1219–1232.
107 Tcherkez G (2016) The mechanism of Rubisco-catalysed oxygenation. *Plant Cell Environ* **39**, 983–997.
108 Luo G, Ono S, Beukes NJ, Wang DT, Xie S and Summons RE (2016) Rapid oxygenation of Earth's atmosphere 2.33 billion years ago. *Sci Adv* **2**, e1600134.
109 He H, Wu X, Xian H, Zhu J, Yang Y, Lv Y, Li Y and Konhauser KO (2021) An abiotic source of Archean hydrogen peroxide and oxygen that pre-dates oxygenic photosynthesis. *Nat Commun* **12**, 6611.
110 He H, Wu X, Zhu J, Lin M, Lv Y, Xian H, Yang Y, Lin X, Li S, Li Y *et al.* (2023) A mineral-based origin of Earth's initial hydrogen peroxide and molecular oxygen. *Proc Natl Acad Sci USA* **120**, e2221984120.
111 Stone J, Edgar JO, Gould JA and Telling J (2022) Tectonically-driven oxidant production in the hot biosphere. *Nat Commun* **13**, 4529.
112 Carpena X, Loprasert S, Mongkolsuk S, Switala J, Loewen PC and Fita I (2003) Catalase-peroxidase KatG of *Burkholderia pseudomallei* at 1.7A resolution. *J Mol Biol* **327**, 475–489.
113 Brioukhanov AL and Netrusov AI (2007) Aerotolerance of strictly anaerobic microorganisms and factors of defense against oxidative stress: a review. *Appl Biochem Microbiol* **43**, 567–582.
114 Harada M, Akiyama A, Furukawa R, Yokobori S, Tajika E and Yamagishi A (2021) Evolution of superoxide dismutases and catalases in cyanobacteria: occurrence of the antioxidant enzyme genes before the rise of atmospheric oxygen. *J Mol Evol* **89**, 527–543.
115 Koppenol WH and Sies H (2024) Was hydrogen peroxide present before the arrival of oxygenic photosynthesis? The important role of iron(II) in the Archean ocean. *Redox Biol* **69**, 103012.
116 Anbar AD, Duan Y, Lyons TW, Arnold GL, Kendall B, Creaser RA, Kaufman AJ, Gordon GW, Scott C, Garvin J *et al.* (2007) A whiff of oxygen before the great oxidation event? *Science* **317**, 1903–1906.
117 Slotznick SP, Johnson JE, Rasmussen B, Raub TD, Webb SM, Zi JW, Kirschvink JL and Fischer WW (2022) Reexamination of 2.5-Ga "whiff" of oxygen interval points to anoxic ocean before GOE. *Sci Adv* **8**, eabj7190.
118 Anbar AD, Buick R, Gordon GW, Johnson AC, Kendall B, Lyons TW, Ostrander CM, Planavsky NJ, Reinhard CT and Stüeken EE (2023) Technical comment on "reexamination of 2.5-Ga 'whiff' of oxygen interval points to anoxic ocean before GOE". *Sci Adv* **9**, eabq3736.
119 Slotznick SP, Johnson JE, Rasmussen B, Raub TD, Webb SM, Zi J-W, Kirschvink JL and Fischer WW (2023) Response to comment on "reexamination of 2.5-Ga 'whiff' of oxygen interval points to anoxic ocean before GOE". *Sci Adv* **9**, eadg1530.
120 Planavsky NJ, Reinhard CT, Wang X, Thomson D, McGoldrick P, Rainbird RH, Johnson T, Fischer WW

and Lyons TW (2014) Low mid-proterozoic atmospheric oxygen levels and the delayed rise of animals. *Science* **346**, 635–638.
121 Zimorski V, Mentel M, Tielens AGM and Martin WF (2019) Energy metabolism in anaerobic eukaryotes and Earth's late oxygenation. *Free Radic Biol Med* **140**, 279–294.
122 Budd GE (2008) The earliest fossil record of the animals and its significance. *Philos Trans R Soc B* **363**, 1425–1434.
123 Brocks JJ, Nettersheim BJ, Adam P, Schaeffer P, Jarrett AJM, Güneli N, Liyanage T, van Maldegem LM, Hallmann C and Hope JM (2023) Lost world of complex life and the late rise of the eukaryotic crown. *Nature* **618**, 767–773.
124 Müller M, Mentel M, van Hellemond JJ, Henze K, Woehle C, Gould SB, Yu R-Y, van der Giezen M, Tielens AGM and Martin WF (2012) Biochemistry and evolution of anaerobic energy metabolism in eukaryotes. *Microbiol Mol Biol Rev* **76**, 444–495.
125 Towe KM (1970) Oxygen collagen priority and early metazoan fossil record. *Proc Natl Acad Sci USA* **65**, 781–788.
126 Harrison JF, Kaiser A and van den Brooks JM (2010) Atmospheric oxygen level and the evolution of insect body size. *Proc R Soc B* **277**, 1937–1946.
127 Gould SB, Garg SG, Handrich M, Nelson-Sathi S, Gruenheit N, Tielens AGM and Martin WF (2019) Adaptation to life on land at high O_2 via transition from ferredoxin-to NADH-dependent redox balance. *Proc Biol Sci* **286**, 20191491.
128 Hu Y and Ribbe MW (2015) Nitrogenase and homologs. *J Biol Inorg Chem* **20**, 435–445.
129 Szenk M, Dill KA and de Graff ARM (2017) Why do fast-growing bacteria enter overflow metabolism? Testing the membrane real estate hypothesis. *Cell Syst* **5**, 95–104.
130 Tran QH and Unden G (1998) Changes in the proton potential and the cellular energetics of *Escherichia coli* during growth by aerobic and anaerobic respiration or by fermentation. *Eur J Biochem* **251**, 538–543.
131 Pfeiffer T and Morley A (2014) An evolutionary perspective on the Crabtree effect. *Front Mol Biosci* **1**, 00017.
132 Han H, Hemp J, Pace LA, Ouyang H, Ganesan K, Roh JH, Daldal F, Blanke SR and Gennis RB (2011) Adaptation of aerobic respiration to low O_2 environments. *Proc Natl Acad Sci USA* **108**, 14109–14114.
133 Arnold BJ, Huang IT and Hanage WP (2021) Horizontal gene transfer and adaptive evolution in bacteria. *Nat Rev Microbiol* **20**, 206–218.
134 Osborne JP and Gennis RB (1999) Sequence analysis of cytochrome bd oxidase suggests a revised topology for subunit I. *Biochim Biophys Acta* **1410**, 32–50.
135 Weiss MC, Sousa FL, Mrnjavac N, Neukirchen S, Roettger M, Nelson-Sathi S and Martin WF (2016) The physiology and habitat of the last universal common ancestor. *Nat Microbiol* **1**, 1–8.
136 Wimmer JLE, Xavier JC, Vieira ADN, Pereira DPH, Leidner J, Sousa FL, Kleinermanns K, Preiner M and Martin WF (2021) Energy at origins: favorable thermodynamics of biosynthetic reactions in the last universal common ancestor (LUCA). *Front Microbiol* **12**, 793664.
137 Jasniewski AJ, Sickerman NS, Hu Y and Ribbe MW (2018) The Fe protein: an unsung hero of nitrogenase. *Inorganics* **6**, 25.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Fig. S1. Correlation of protein family verticality and frequency among prokaryotic genomes.
Fig. S2. Correlation of reaction frequency versus genome size.
Fig. S3. Distribution of KEGG functional categories within prokaryotic phyla associated with protein families catalyzing oxygen-dependent reactions.
Fig. S4. Taxonomic distribution of O_2-dependent and O_2-independent reactions.
Fig. S5. Reaction frequency versus genome size for O_2-dependent reactions in prokaryotes.
Fig. S6. Distribution of Gibbs energy ΔG for oxygen-dependent and oxygen-independent reactions.
Fig. S7. Alternative plots of the average verticality across the 10 functional categories with the highest frequency of O_2-utilizing protein families.
Fig. S8. Origins of O_2-dependent reactions across a backbone phylogeny.
Table S1. List of O_2-dependent and O_2-independent reactions from KEGG that were linked to protein families.
Table S2. All protein families with respective KEGG Orthology identifier (KO), corresponding reactions, protein family size and verticality V.
Table S3. Statistical tests with relevant parameters.
Table S4. Gibbs energy ΔG for O_2-dependent and -independent reactions.
Table S5. Most common reactants and products across 365 O_2-dependent reactions of prokaryotes.
Table S6. Most frequent H_2O_2 – O_2 interconverting enzymes among 365 O_2-dependent reactions.
Table S7. Cofactors for 365 O_2-dependent reactions.
Table S8. O_2-dependent reactions per genome across taxonomic groups.

Table S9. Information on the 5655 prokaryotic genomes used in this study.
Table S10. Enzyme commission numbers and reaction type for 365 O_2-dependent reactions.
Table S11. Ancestral state reconstruction for 365 O_2-dependent reactions.
Table S12. Functional categories with a list of reactions they include, number of protein families per category and their respective average verticality.