Geobiological feedbacks, oxygen, and the evolution of nitrogenase

Authors: Florence Mus, Daniel R. Colman, John W. Peters, and Eric S. Boyd

Mus, Florence, Daniel R. Colman, John W. Peters, and Eric S. Boyd. "Geobiological feedbacks, oxygen, and the evolution of nitrogenase." Free Radical Biology & Medicine (February 2019). DOI:10.1016/j.freeradbiomed.2019.01.050.

Accepted Manuscript

Geobiological feedbacks, oxygen, and the evolution of nitrogenase

Florence Mus, Daniel R. Colman, John W. Peters, Eric S. Boyd

PII: S0891-5849(18)32253-6

DOI: https://doi.org/10.1016/j.freeradbiomed.2019.01.050

Reference: FRB 14148

To appear in: *Free Radical Biology and Medicine*

Received Date: 23 October 2018

Accepted Date: 31 January 2019

Please cite this article as: F. Mus, D.R. Colman, J.W. Peters, E.S. Boyd, Geobiological feedbacks, oxygen, and the evolution of nitrogenase, *Free Radical Biology and Medicine* (2019), doi: https://doi.org/10.1016/j.freeradbiomed.2019.01.050.

Geobiological Feedbacks, Oxygen, and the Evolution of Nitrogenase

Florence Mus[1],‡ Daniel R. Colman[2],‡ John W. Peters[1*] and Eric S. Boyd[2*]

[1]Institute of Biological Chemistry, Washington State University, Pullman, WA

[2]Department of Microbiology and Immunology and the NASA Astrobiology Institute, Montana State University, Bozeman, MT

Key words: Biological Nitrogen fixation, nitrogenase, molybdenum, vanadium, iron, Great Oxidation Event, oxygen, methanogens, iron-sulfur

‡ Both authors contributed equally to this work

*Address correspondence to John W. Peters, (jw.peters@wsu.edu) and Eric S. Boyd, (eric.boyd@montana.edu)

Abstract

Biological nitrogen fixation via the activity of nitrogenase is arguably one of the most important biological innovations, allowing for an increase in global productivity that eventually permitted the emergence of complex forms of life. The complex metalloenzyme termed nitrogenase contains complex iron-sulfur cofactors. Three versions of nitrogenase exist that differ mainly by the presence or absence of a heterometal at the active site metal cluster (either Mo or V). Mo-dependent nitrogenase is the most common while V-dependent or heterometal independent (Fe-only) are often termed alternative nitrogenases since they have lower activities and are expressed in the absence of Mo. Phylogenetic evidence indicates that biological nitrogen fixation emerged in an anaerobic, thermophilic ancestor of hydrogenotrophic methanogens and later diversified via lateral gene transfer into anaerobic bacteria, and eventually aerobic bacteria. Isotopic evidence indicates that nitrogenase activity existed at 3.2 Ga prior to the advent of oxygenic photosynthesis and rise of oxygen in the atmosphere, implying the presence of favorable environmental conditions for oxygen-sensitive nitrogenase to evolve. Following the proliferation of oxygenic phototrophs, diazotrophic organisms had to develop strategies to protect nitrogenase from oxygen inactivation and generate the right balance of low potential reducing equivalents and cellular energy for growth and nitrogen fixation activity. Here we review the fundamental advances in our understanding of biological nitrogen fixation in the context of the emergence, evolution, and distribution of nitrogenase, with an emphasis placed on key events associated with its emergence and diversification from anoxic into oxic environmental conditions.

1. Introduction

Fixed nitrogen (N) in the form of ammonia (NH_3) is essential for all forms of life where it is used in the synthesis of basic building blocks of biomolecules such as DNA, RNA, and protein. Atmospheric nitrogen (N_2) is readily abundant but essentially inert due to the high activation energy associated with the triple bond between nitrogen atoms. Fixed nitrogen is formed through the conversion of N_2 to NH_3, ammonium ion (NH_4^+), or nitrogen oxides such as nitrate (NO_3^-) or nitrite (NO_2^-). Several abiotic processes can overcome the activation barrier associated with the N_2 triple bond, namely lightning-based oxidative atmospheric processes and mineral-based reductive processes, which are discussed below. The combined fixed N input of these processes, however, is thought to have been insufficient to meet the fixed N demand of an increasingly productive biosphere [1], [2], thereby providing the impetus to evolve and refine biological mechanisms to drive the reduction of N_2 [3], [4].

Biological nitrogen fixation refers to the enzyme-catalyzed reduction of N_2 to NH_3 and occurs exclusively in microorganisms called diazotrophs. Diazotrophic activity accounts for approximately two-thirds of the fixed N available to the biosphere [5], and these organisms are distributed across multiple phyla of Bacteria and Archaea [6]. As discussed in more detail below, diazotrophs represent ~7.8% of all bacterial and archaeal cultivars with sequenced genomes and are represented by organisms operating aerobic, anaerobic, phototrophic, chemotrophic, autotrophic and heterotrophic metabolisms [7]. This implies that a minority of organisms (at a taxonomic level) have evolved to supply the fixed N demand of numerous co-inhabiting taxa. Despite nitrogenase being extremely sensitive to oxygen, these data suggest that diazotrophs have acquired mechanisms to supply fixed N to organisms inhabiting a broad range of oxic and anoxic surface and subsurface ecosystems.

Contemporary nitrogenases are amongst the most, if not the most, complex metalloenzymes in biology [8]. Nitrogenase exists in molybdenum (Mo)-dependent (Nif), vanadium (V)-dependent (Vnf), and iron (Fe)-dependent or heterometal-independent (Anf) forms [9]. Nif are by far the most well-characterized and most commonly occurring forms of nitrogenase [4],[7]. Nif is made up of two separable protein components termed the Fe protein and MoFe protein to reflect the relative composition of their associated metal-containing prosthetic groups (Fig. 1A, 1B). The Fe protein serves as the unique electron donor to the MoFe protein that harbors the active site metallocluster where nitrogen is reduced to ammonia [8]. The Fe protein, encoded by *nifH*, is a dimer of identical subunits bridged by a single [4Fe-4S] cluster and is a member of a large protein family that couples the binding and hydrolysis of nucleotide triphosphates to conformation changes [8]. The MoFe protein contains two complex metal centers, one of which is termed the P cluster (8Fe-7S) that serves as an intermediate electron carrier, and the other which is termed the FeMo-cofactor (Mo-7Fe having homocitrate and carbide non-protein ligands) that serves as the substrate reduction site [10]. The MoFe protein is a $\alpha_2\beta_2$ heterotetramer encoded by the *nifD* and *nifK* genes where the two P clusters are bridged by alpha (NifD) and beta (NifK) subunits and the two FeMo-cofactors are contained within the alpha (NifD) subunits [8]. During nitrogenase catalysis the Fe protein and MoFe protein associate and dissociate in a cycle involving the coupling of the hydrolysis of two MgATP with the transfer of an electron from the Fe protein to the MoFe protein. While the conversion of N_2 to NH_3 is a 6 electron reduction (3 electrons for each of the two NH_3 molecules), the minimal reaction stoichiometry for this reaction requires eight electrons, eight protons, and sixteen MgATP since for each N_2 reduced there is concomitant production of one molecule of H_2 (Fig. 1A, 1B).

Vnf (encoded by *nifDK* homologs *vnfDK* and a small non-homologous subunit *vnfG*) and heterometal independent or Anf (encoded by *nifDK* homologs *anfDKG*) nitrogenases only occur in a subset of nitrogen fixing organisms that also encode Nif [3], [4], [7]. In these organisms and under laboratory conditions, alternative Anf or Vnf are expressed under Mo limitation [11], [12]. More recent evidence indicates that controls on the expression of nitrogenase in organisms inhabiting natural environments might be more complex than previously recognized, and may involve temperature, salinity, as well as trace element availabilities [13], [14], [15].

There has been significant interest in the timeline for the evolutionary origin of biological nitrogen fixation (e.g., [16]), since the availability of fixed N often limits productivity in ecosystems today and is likely to have done so since early in Earth history. Moreover, while the exact time in Earth history when this keystone process first originated cannot be known definitively, various lines of evidence can be brought together to constrain the timing for nitrogenase emergence. In particular, the evolution of enzymes and proteins that form the structural components of nitrogenase and that synthesize the active site metallocluster where N_2 reduction occurs can be placed in relative time with paralogous proteins and enzymes involved in various aspects of methanogenesis and photosynthesis using phylogenetic approaches (e.g., [4], [4]). Such information can then be cross-referenced with isotopic evidence for biological N_2 fixation in the rock record and paleostratigraphic data describing the availability of key trace elements required for nitrogenase through geological time to constrain the timeframe when key events in the evolution of this keystone process took place. This review will address the current state of knowledge on the origin and the evolutionary history of biological nitrogen fixation. The influence of oxygen on the evolution of Nif during its diversification from anaerobic to aerobic

metabolic backgrounds will be discussed with an emphasis on the advent of unique life styles and physiological mechanisms that allow for N_2 fixation in otherwise oxygen-rich environments.

2. Abundance of fixed nitrogen on early earth

Nitrogen is essential for life on Earth where it is an integral ingredient in the synthesis of amino acids, nucleic acids, and other biological molecules. Most of the nitrogen on Earth is in the form of N_2, which is generally considered to be chemically inert and bio-unavailable. Nonetheless, early models for the composition of the early atmosphere suggest that it was reduced and had abundant supplies of ammonia produced by equilibration of N_2 with minerals in a yet to be differentiated Earth (e.g., [17]). However, more current models suggest that the early Earth atmosphere was mildly oxidizing and that ammonia was likely in limited supply due to inescapable photolytic reactions involving this molecule [2]. Several abiotic processes have been described as potential sources of the limited amounts of fixed N that must have been present in aqueous environments to fuel life on early Earth, including lightning-based oxidative atmospheric processes with subsequent deposition of oxidized nitrogen compounds to surface environments and subsurface mineral-based reductive processes that provide reduced forms of fixed N such as ammonia (e.g., [18], [19]). The flux of fixed N supplied by abiotic mechanisms on early Earth must have been finite and this, when combined with an increased fixed N demand from an ever-expanding global microbiome, led to fixed N limitation which may have precipitated the innovation of biological mechanisms to reduce N_2 [3], [4].

3. Emergence of biological nitrogen fixation

The metabolic costs associated with fixing N_2 (minimally 16 ATPs, 8 reducing equivalents per N_2 reduced) may lead one to imagine that this process originated in a N limited environment and in an ancestral organism operating a high energy yielding metabolism such as photosynthesis or aerobic respiration. Paradoxically, however, phylogenetic reconstruction of NifHDK structural proteins (Fig. 1B) indicates that the earliest evolving organisms capable of N_2 fixation were ancestors of thermophilic hydrogenotrophic methanogens [4], [20], [7], implying an origin in an anoxic environment. Since Nif functions to relieve fixed N limitation, these collective observations suggest that Nif originated in a high-temperature, anoxic environment that was limited in available N. Energetically, hydrogenotrophic methanogenesis is one of the lowest energy yielding reactions capable of sustaining life. For these reasons, it seems contradictory that selective pressure to evolve the energy intensive process of N_2 fixation would act on an ancestor of hydrogenotrophic methanogens. We previously outlined a rationale for an anaerobic, hydrogenotrophic, and thermophilic methanogen Nif ancestor [21] and briefly review this here.

If we assume life Nif was not a property of the first forms of life on Earth, then that life would have been dependent on the availability of fixed sources of N (i.e., NO_3^-, NH_3/NH_4^+) supplied from its geological environment. Thus, the emergence of Nif would have had to occur in an environment in which the depletion of abiotic fixed nitrogen resulted in selective pressure to evolve Nif. Lightning can break the triple bond of N_2, allowing it to combine with other atmospheric gases and produce compounds such as nitrous oxide that can be deposited on land and in oceans [18], [2], [1]. However, there is substantial debate regarding the amount of fixed N that could be supplied by abiotic atmospheric processes early in Earth history, with most models predicting low fluxes [2]. Moreover, this flux would have been localized to surface

environments, leaving organisms inhabiting deep sea marine environments (e.g., methanogens in hydrothermal environments) limited in fixed N due to preferential use in the water column. A second source of fixed N possibly supporting methanogens living near hydrothermal vents may have included NH_3 generated through cooling and re-equilibration of volcanic gases and reduction of N_2 via mineral-based catalysts such as reduced iron phases [22], [19]. However, like atmospheric sources of fixed N, the flux of NH_3 from these abiotic sources would have been finite. As the dominant primary producers in many anoxic subsurface environments, methanogens would likely be the first organisms to encounter fixed N limitation. Thus, as anaerobic ecosystems founded on autotrophic methanogens became more productive, demand for fixed N may have outweighed supply, thus, providing the impetus to evolve Nif.

A second consideration in the emergence of Nif in methanogens is that selection acts on existing functional diversity and the observations that Nif is evolutionarily related to enzymes that function in the biosynthesis of cofactor F_{430}. The nickel-containing hydroporphinoid F_{430} is a key cofactor in the metabolism of methanogens, archaeal methanotrophs, and several recently evolved bacteria. The involvement of several nitrogenase protein paralogs in the biosynthesis of cofactor F_{430} was predicted [23] and recently confirmed [24], [25], [26] (Fig. 2). These include paralogs of two of the three structural components of Nif, NifHD, which were termed Nif-like proteins, Nfl or NflHD [3]. Phylogenetic reconstructions of NifHD and NflHD indicate that NflHD predate NifHD [4]. The fact that the NflHD complex lacks a paralog of NifK, combined with the observation that NifHD is derived from an ancestor of NflHD, suggests that the duplication of an ancestor of NifD to generate the ancestor of NifK took place after the divergence of NflHD and NifHD from their ancestor [6] (Fig. 2). Given the limited distribution

of Nfl in methanogens and related archaeal methanotrophs, it would be surprising if Nif evolved in any organism other than in an ancestor of these organisms.

An additional consideration involves the stability of N_2, which is often considered inert because of the high activation energy associated with its ternary N≡N triple bond (~940 kJ mol^{-1}). To overcome the high activation energy for N_2 reduction and favor formation of products, the reaction can be operated at high temperature and at elevated pressure in the presence of high concentrations of reactants and iron-based catalysts in what is known as the Haber-Bosch process. To maintain the reaction favoring product formation, ammonia must be continuously removed. This would be easy to accomplish via mixing of hydrothermal vent fluids with ocean water. Intriguingly, at elevated but still biologically relevant temperatures (<121°C) and in the presence of elevated pressure and high concentrations of reactants, the equilibrium for the reaction should also permit product formation if appropriate catalysts are available. Indeed, experiments conducted in the presence of Fe-S minerals under conditions similar to those present at hydrothermal vents were shown to promote reduction of N_2 to ammonia, albeit with meager yields [22]. Such conditions may have provided the impetus for an H_2-dependent methanogen to evolve a proto nitrogenase with an Fe-S containing active site to overcome the activation barrier for N_2 reduction, push the reaction toward product formation, and relieve NH_3 limitation. Selection and refinement of this primitive enzyme complex over time would allow for further improvements in its efficiency, eventually allowing the process to diversify into lower-temperature and lower-pressure environments.

4. When Did Nitrogenase Emerge?

Reports of the time frame when nitrogenase emerged are variable and are dependent on the analytical method and approach used to make the estimation. Isotopic analyses of nitrogen preserved in kerogens reveal values that are consistent with input from atmospheric nitrogen, presumably biological nitrogen fixation, prior to 2.5 Ga [27], [28], [29]. More recent analyses of marine and fluvial sediment rocks dated to nearly 3.2 Ga reveal values that are suggestive of the presence of biological nitrogen fixation at this time [30]. Thus, from an isotopic perspective, it is possible that biological nitrogen fixation was in place by 3.2 Ga. Perhaps consistent with this time frame, a recent phylogenetic analysis placed Nif in the Last Universal Common Ancestor (LUCA) of Archaea and Bacteria [31]. Given the limited distribution of Nif gene complements among Archaea (restricted to several lineages of methanogens, as described below) and the nesting of bacterial NifHDK sequences among these archaeal NifHDK sequences [6], the previous interpretation [31] of Nif being a property of the LUCA of Archaea and Bacteria based on phylogenetic reasoning is almost certainly incorrect.

If Nif was responsible for biological nitrogen fixation very early in Earth history then it appears that it differed at least in metal-cofactor composition from what is present in extant organisms. In model diazotrophs that have been studied physiologically and biochemically, active nitrogenase expression requires a small suite of accessory genes. The genes required include a S-adenosyl methionine dependent enzyme (NifB) that modifies an iron-sulfur precursor by the addition of a carbide that simultaneously coordinates 6 Fe ions in a (6Fe-6S) complex (NifB-co) and a scaffold on which the FeMo-co is built that evolved through an in tandem gene duplication of NifDK (NifEN) (Fig. 3A, 3B). Phylogenetic dating of key evolutionary events (e.g., duplication of the ancestor of NifDK to NifEN) in the evolutionary history of nitrogenase suggests that Nif emerged ~ 2.1 Ga [4] (Fig. 4). How can these two (i.e., 3.2 versus 2.1 Ga)

seemingly contradictory time frames for the emergence of biological nitrogen fixation be reconciled?

The duplication of an ancestor of NifDK to NifEN (Fig. 3) apparently resulted in the ability to further mature NifB-co to FeMo-co, an evolutionary event that would have likely increased the specificity and activity of the ancestral enzyme where these clusters were bound. However, as mentioned above, phylogenetic dating approaches suggest that this event did not occur until ~ 2.1 Ga (Fig. 4), thereby leaving a gap of 1.1 Ga between when isotopic evidence (3.2 Ga; [30]) indicates N_2 fixation was in place and when phylogenetic evidence (2.1 Ga; [6]) indicates this process in its contemporary form (e.g., with FeMo-co at the active site of Nif) emerged. Perhaps reconciling this discrepancy is the observation that the duplication of the ancestor of *nifD* resulting in *nifK* took place as early as 3.5 Ga [4], thereby preceding the duplication of the ancestor of *nifDK* to *nifEN* by ~ 1.4 Ga (Fig. 4). Thus, it is possible that a primitive NifDK protein harboring an active site metal cluster similar to NifB-co [32] could be responsible for N isotopic signatures in ancient marine and fluvial sediments at 3.2 Ga. As opposed to FeMo-co that requires an elaborate biosynthetic pathway involving numerous proteins to assemble, the synthesis of NifB-co requires only ferrous iron and sulfide, which self-assemble into molecular clusters, and the activity of NifB [33]. NifB is a member of the radical SAM protein superfamily, one of the oldest protein families in biology [34], [35], [36]. Thus, it is possible that an early evolving methanogen that encoded for NifHDKB adapted iron sulfur molecular clusters that are common in hydrothermal vent plumes and other reduced environments [37], [38] to synthesize a NifB-co like cluster that supported nitrogenase activity.

If a proto-nitrogenase emerged ~ 3.5 Ga, then what was the evolutionary trigger to evolve a catalytically more efficient Mo dependent nitrogenase (Nif) at ~ 2.1 Ga? It has been suggested

that the trigger for the emergence of Nif from an alternative form of nitrogenase containing iron (Anf) (Fig. 5B) was the increased mobility and bioavailability of Mo that accompanied the widespread oxygenation of the biosphere beginning at ~ 2.45 Ga [39], [40]. However, phylogenetic reconstructions indicate that Anf is derived from Nif [6], [20] (Fig. 5A), a finding that is consistent with the observation that all Anf-encoding genomes also encode Nif [20], [3], [7], and that the biosynthesis of the active site metallocluster of alternative nitrogenases is dependent on Nif-encoded machinery [41], [42], [12]. Thus, Anf is unlikely to have given rise to Nif [4] (Fig. 5A, Fig. 5B), which contradicts what the geochemical record of trace metal availabilities might suggest [39], [40]. However, as discussed above, the enzyme that existed from 3.5 to 2.1 Ga (i.e., prior to the duplication of NifDK to give rise to NifEN) (Fig. 4) may have harbored an active site cluster that comprised iron and sulfide only (i.e., NifB-co). Thus, while the predecessor to Nif is unlikely to have been Anf, the proto-nitrogenase that predated Nif is likely to have been Mo-independent, which would satisfy geochemical arguments for the apparent lack of bioavailable Mo during this time.

A major shift in the global nitrogen cycle was also taking place at the time that the emergence of Nif is thought to have taken place due to the input of oxygen accompanying the proliferation of Cyanobacteria [43]. The increased productivity of ecosystems due to oxygenic photosynthesis would have increased demand for fixed forms of N. At the same time, increased oxygen would provide a mechanism to weather molybdenum sulfides, releasing soluble molybdenum and increasing its bioavailability in the biosphere [44]. However, at the same time that demand for fixed N was increasing for biosynthetic purposes and that Mo was becoming more available [40], which has traditionally been considered to have been the driver of the emergence of Nif [39], organisms that use NH_3 as an electron donor through the process of

nitrification emerged [28], [29]. Nitrification, or the stepwise oxidation of NH_3 to nitrite and ultimately nitrate, requires O_2 and is thus thought to not have evolved prior to the buildup of oxygen ~ 2.5 Ga (Fig. 4). Indeed, keragens dated to ~ 2.5 Ga host isotopically heavy nitrogen, indicating that loss of isotopically light N through a coupled nitrification and denitrification cycle was likely taking place at this time [28], [29]. This observation implies that both NH_3 and O_2 must have been available in sufficient supply, at least locally, to select for aerobic nitrifying organisms.

Two important points can be gleaned from this observation. Firstly, if NH_3 was available to select for nitrifying organisms at this time, as suggested above, then biological demand for NH_3 for biosynthetic purposes up until this time was not necessarily outweighing the supply. Again, this supply could have been abiotic in the form of lightning-based oxidation of N_2 or mineral-based reduction of N_2 or biotic in the form of a proto-NifHDK enzyme perhaps with an active site cofactor resembling NifB-co, at least based on phylogenetic evidence [4], [32]. Secondly, nitrification coupled with denitrification results in a net loss (as N_2) of fixed N from an environment. Thus, the emergence of a coupled nitrification-denitrification cycle at ~2.5 Ga (Fig. 4) would represent a new and potentially substantial sink for fixed N that may have tilted the balance between the supply and the demand for fixed N. This in turn, may have represented the selective pressure for the duplication of an ancestor of *nifDK* to *nifEN*, allowing for the further maturation of NifB-co to FeMo-co and the evolution of a superior nitrogenase in the form of Nif. The estimated timing of the emergence of a coupled nitrification/denitrification cycle at no later than 2.5 Ga via isotopic approaches [28], [29] (Fig. 4) and an estimated timing of the duplication of an ancestor of *nifDK* to yield *nifEN* at ~2.1 Ga via phylogenetic approaches [20] is consistent with this hypothesis.

5. Taxonomic distribution of Nif

At a taxonomic level, the compliment of the minimum set of genes encoding a functional Nif (i.e., NifHDKEB; [6], [45], [46]) (Fig. 1A, 1B) exhibit sporadic distributions among the genomes of Bacteria and Archaea. For example, an analysis of genomes available in public databases in 2010 revealed a limited distribution of Nif among Archaea, with the only identified homologs in methanogenic Euryarchaeota [4], [6], [20]. In contrast, these same studies found that Nif was broadly distributed among members of the Bacteria. The patterns in the taxonomic distribution of Nif were interpreted to result from numerous instances of lateral gene transfer and/or gene loss events involving Archaea and Bacteria (Fig. 6A, 6B), with an origin for Nif among methanogenic Archaea, as described above. A more recent survey of the distribution of Nif among available genomes confirmed this taxonomic distribution [7].

Since the genomes surveyed for the distribution of Nif previously were largely from cultivars and to account for the possibility that the limited diversity available in culture databases obfuscates an accurate depiction of the taxonomic distribution of Nif, we evaluated an additional 4,730 and 791 publicly available bacterial and archaeal genomes for the presence of genes encoding Nif. Many of these genomes were assembled from environmental metagenomes and represent lineages without previously confirmed evidence for nitrogenase genes. Through this effort, an additional five bacterial taxonomic divisions (i.e., Elusimicrobia, Planctomycetes, Lentisphaerae, Fibrobacteres, and the candidate division ‘Margulisbacteria’; Fig. 6B) were identified with at least one representative encoding the genetic potential for nitrogen fixation via Nif. Likewise, evidence for genes encoding Nif were identified in the euryarchaeal order Methanomasiliicoccales (Fig. 6A). Consistent with previous studies indicating that the

distribution of Nif is limited to methanogenic Archaea, members of the Methanomasiliicoccales have thus far only known to be methanogenic [47]. Consequently, while the presence of nitrogenase functionality appears to be widely distributed among Bacteria from numerous taxonomic divisions that exhibit diverse physiological strategies [6], [20], [7], the capacity for nitrogen fixation appears restricted to several lineages of methanogenic Euryarchaeota.

Using a previously curated database of 359 organisms with available genomes that encode for the minimum complement of Nif genes (46 of which have been experimentally shown to fix N_2), we defined the distribution of Nif as a function of the primary mode of energy metabolism (phototroph versus chemotroph and aerobe, anaerobe, or facultative anaerobe). Of the 359 diazotrophic organisms with available genomes identified previously, 66 (18%) are phototrophic taxa and 293 (82%) are chemotrophic taxa [7]. Among the 66 phototrophic diazotrophs, 19 are aerobes (Cyanobacteria), 21 are anaerobes (Chlorobi, Chloroflexi, Firmicutes, and Proteobacteria), and 24 are facultatively anaerobes (Chloroflexi, Cyanobacteria, and Proteobacteria). Among the 293 chemotrophic diazotrophs, 79 are aerobes (Actinobacteria, Nitrospirae, Proteobacteria, and Verrucumicrobia), 150 are anaerobes (Bacteroidetes, Deferribacteres, Euryarchaeota, Firmicutes, Fusobacteria, Proteobacteria, Spirochaetes, Thermodesulfobacteria), and 63 are facultatively anaerobes (Firmicutes, Proteobacteria, Verrucumicrobia). Several of the organisms could not be classified with respect to their mode of energy metabolism.

6. Early diazotrophs in anoxic and oxic worlds

As reviewed above, phylogenetic evidence indicates that nitrogenase emerged and diversified in hydrogenotrophic methanogens (Archaea), implying an origin in an anoxic environment [4], [6], [45], [7]. An anoxic origin for Nif is consistent with the oxygen sensitivity

of this enzyme [8]. Phylogenetic analyses indicate that an anaerobic bacterium affiliated with the Firmicutes likely obtained Nif via lateral gene transfer from an ancestral methanogen [4], [6]. An origin and ensuing proliferation of Nif would likely have occurred in a fixed N limited environment where anaerobic methanogens and Firmicutes coexisted and where Mo was at least episodically available, such as in a redox stratified Proterozoic ocean basin [4]. Alternatively, given that the solubility of Mo even under anoxic and sulfidic conditions is enhanced under alkaline conditions [48], [49], [50], [51], it is possible that diversification of Nif into Firmicutes and other Bacteria may have taken place in a hydrothermal or mesothermal hydrothermal vent system similar to Lost City [52] or Strytan, Iceland [53]. The ensuing proliferation of Nif among anaerobic bacteria involved acquisition of the minimum complement of genes required to synthesize an active Nif (i.e., NifHDKEB) in anaerobic Proteobacteria and Chlorobi, among others, and ultimately into lineages with the ability to integrate O_2 into their energy metabolism [45]. Intriguingly, once Nif was acquired in aerobic lineages, it was apparently never laterally transferred back to lineages characterized by strictly anaerobic members, resulting in a stark delineation of aerobic and anaerobic diazotrophs in phylogenetic reconstructions of Nif [45], [21]. While we did not explicitly quantify when Nif diversified into aerobic metabolic backgrounds in our previous work [6], it is perhaps reasonable to assume that this accompanied the second major rise in the availability of O_2 that took place ~ 800 Ma [54] (Fig. 4).

Regardless of when Nif was acquired among aerobic Bacteria, it is apparent that this was accompanied by substantial recruitment and loss of *nif* genes [45], as well as a reconfiguration of mechanisms of electron transport to Nif [7]. At a broad level, the number of *nif* encoded genes increased substantially during the diversification of Nif encoding organisms, with early evolving anaerobic lineages (e.g., methanogens, clostridia) tend to harbor less complex suites of *nif* genes

than more recently evolved aerobic lineages (e.g., *Proteobacteria*, *Cyanobacteria*). For example, the number of genes in anaerobic diazotrophic *nif* gene clusters (minimum of 5; i.e., NifHDKEB) increased substantially to nearly 20 in the most derived aerobic diazotroph (e.g., *Azotobacter vinelandii*). The putative roles of these accessory gene products suggest that many of them (i.e., NifZ, NifW, NifQ, NifT) were recruited to improve the maturation of FeMo-co or MoFe metal clusters (Fig. 7). While the observed increase in the number of *nif* genes (e.g., NifL, NifA, Rnf, and Fix) and their phylogenetic distribution are strongly correlated with adaptation to integrate oxygen into their energy (Fig. 7), the increase is not correlated with any of the known oxygen protection mechanisms [45]. Rather, the majority of genes recruited or lost during the transition of nitrogen fixation from anaerobes to aerobes appear to be associated with metal cluster biosynthesis and nitrogenase regulation. Consistent with this hypothesis, the transition of Nif from anoxic to oxic environments is associated with a shift from $NifI_1$-I_2-dependent posttranslational regulation in anaerobes to NifA-dependent transcriptional regulation in obligate aerobes and facultative anaerobes (Fig. 7). In this context, both the increase in the number of *nif* genes and the shift in their mode of regulation during the transition of Nif from anaerobic to aerobic metabolism appear to be due to selection for increased efficiency in the synthesis of Nif to meet the higher fixed N demands associated with more productive aerobic metabolism and to more efficiently regulate Nif under oxic conditions that favor protein turnover. Given that fixed nitrogen often limits ecosystem productivity, our previous observations indicating a major shift in the composition and function of *nif* regulons [45] further underscore the dynamic interplay between the evolution of Earth's oxygen, nitrogen, and carbon biogeochemical cycles.

7. Mechanisms for preventing oxygen inactivation of nitrogenase

N_2 fixation is associated with a diversity of microorganisms that display a wide variety of physiologies that range from obligate anaerobes to obligate aerobes [3], [6], [46]. A number of different strategies of dealing with oxygen have been identified among diazotrophs [55], [56]. The simplest strategy to prevent oxygen-dependent inactivation of nitrogenase is for cells to only fix nitrogen in anoxic environments or during anaerobic metabolism. Strict anaerobes, such as *Clostridium* spp., *Desulphovibrio* spp. and photosynthetic sulfur bacteria such as *Chlorobium* and *Chromatium*, cannot grow in the presence of oxygen regardless of whether they are fixing nitrogen [56]. Facultative anaerobes such as *Klebsiella pneumoniae* or *Rhodobacter capsulatus* that, while capable of aerobic growth, temporally separate nitrogen fixation from aerobic metabolism and fix nitrogen only when growing anaerobically or microaerobically [56].

Unicellular Cyanobacteria also only fix nitrogen anaerobically in environments such as microbial mat communities [57]. Cyanobacteria in these communities conduct oxygenic photosynthesis during the day resulting in CO_2 fixation and the storage of carbon in the form of starch during the daylight hours [58]. At night the respiratory capacity of the concentrated mat biomass can quickly consume the available oxygen resulting in an anoxic environment. Cyanobacteria, in turn, transition to an anaerobic mode of metabolism and fix nitrogen with energy provided through the fermentation of stored starch [59]. In contrast, filamentous Cyanobacteria can protect nitrogenase from inactivation by exposure to oxygen through spatially separating nitrogen fixation from oxygenic photosynthesis by conducting nitrogen fixation in specialized heterocyst cells where oxygenic photosynthesis does not occur [60]. These heterocysts are provided carbon for energy by vegetative cells in the filament and are maintained in an anoxic environment. In turn, nitrogen fixation within the heterocysts provides fixed nitrogen for the growth of vegetative cells. A similar strategy is employed by symbiotic nitrogen

fixers that produced fixed nitrogen only in association with leguminous plants. Legume nodulation is a complex process of signaling between *Rhizobia* and plants resulting in the infection of root hairs and the establishment of root nodules [61]. Plants produce an oxygen binding molecule similar to hemoglobin in animals called leghemoglobin. Leghemoglobin in the nodules binds oxygen creating a microaerobic niche within the nodule at a low enough oxygen tension to prevent the inactivation of nitrogenase but at the same time making oxygen available for oxygen respiration [62].

Due to the energy demand of nitrogen fixation both in the form of the ATP and low potential electrons required, the evolution of mechanisms that allow for nitrogen fixation during oxygen-dependent respiration was a significantly impactful innovation. Obligate aerobes have evolved mechanisms to exploit respiration itself as a mechanism to prevent oxygen inactivation of nitrogenase [63]. Nitrogen fixing members of the *Pseudomonacia* use respiratory protection as a primary mechanism of oxygen protection. Nitrogen fixing *Azotobacter vinelandii* evolved mechanisms to carefully control the consumption of oxygen at the membrane and to control the energy balance of the cells such that nitrogen can be fixed optimally while maintaining an oxygen free reducing environment with the cell cytoplasm [64], [65]. This strategy, when coupled with the production of viscous carbohydrate polymers that promote cellular aggregation [66], provide an effective means to exploit energy rich respiratory metabolism while fixing N_2, resulting in robust diazotrophic activity.

8. Sources of Reducing Equivalents During Aerobic and Anaerobic Diazotrophic Growth

The physiological electron donors to nitrogenase are ferredoxin (Fd) [67], [68], [69], [70] and flavodoxin (Fld) [71], [72] (Fig. 1). Reduction of Fd or Fld in anaerobic diazotrophs can

occur via the oxidation of pyruvate via the activity of pyruvate flavodoxin oxidoreductase (PFOR) [73], [74], [75], [69] or the oxidation of hydrogen through the activity of one of several [FeFe]- or [NiFe]-hydrogenases [76], [77], [78], [79], [80], [81], [82]. However, the reduction of Fd ($E^{o'}$ ~ -420 mV) in aerobic and some anoxygenic phototrophic diazotrophs that inhabit less reducing environments is more of a challenge since these metabolisms typically generate reduced NADH or NADPH ($E^{o'}$ = -320 mV) as the primary electron carrier [7]. These electron carriers are not of low enough potential to drive N_2 reduction.

Anaerobic purple sulfur and facultatively anaerobic non-sulfur anoxygenic phototrophic bacteria utilize photosystems to drive cyclic electron transfer, with reduction of NAD^+/$NADP^+$ accomplished with electrons supplied by oxidation of an exogenous substrate with additional energy from reverse electron transport if the substrate is not of low enough potential to reduce NAD^+/$NADP^+$ [83], [84]. To help overcome the limited ability to generated reduced Fd/Fld for N_2 fixation, these taxa have been proposed to have acquired the Fix complex and/or the *Rhodobacter* nitrogen fixation (Rnf) complex [85]. Fix catalyzes the oxidation of NADH coupled to the reduction of quinone and Fd whereas Rnf harnesses the free energy of the proton/sodium motif force to catalyze the NADH-dependent reduction of Fd [86]. Indeed, genes encoding fix were among those that were shown to be recruited to *nif* gene clusters during the transition of Nif from anaerobic to aerobic backgrounds [45].

Oxygenic phototrophic cyanobacteria can utilize photosystem I to energize electrons to negative enough potentials to drive reduction of Fd [84], [83]. However, this Fd is not available to Nif since, as described above, it must be temporally or spatially separated from oxygenic photosynthesis. Cyanobacteria encode ferredoxin-$NADP^+$ oxidoreductase (FNR) that can function in reverse to reduce Fd or Fld with NADPH generated by carbohydrate oxidation in

heterocysts or when O_2 tensions are low [87], [88], [89], [73]. Some cyanobacteria also encode for PFOR [7]. Like cyanobacteria, anaerobic anoxygenic green sulfur bacteria use a Type I photosystem that is distantly related to photosystem I to generate reduced Fd as a component of photosystem driven cyclic electron transfer [83], [84]. Green sulfur bacteria also encode an FNR that is structurally unrelated to conventional FNR and this can also be used to drive reduction of Fd [7]. Thus, at least seven enzyme complexes have evolved to provide reduced Fd for N_2 fixation: PFOR, [NiFe]-hydrogenase, [FeFe]-hydrogenase, Rnf, Fix, and both forms of FNR [73].

Different Fds and Flds are also likely to be involved in delivery of electrons to nitrogenase in aerobes, anaerobes, and phototrophs, and may vary alongside the primary enzyme that is involved in reducing these electron carriers. Fds are sensitive to O_2 due to the lability of their iron sulfur (FeS) clusters [7]. In contrast, Flds contain flavin mono-nucleotide as the prosthetic group involved in electron transfers instead of FeS clusters and hence are thought to be less sensitive to O_2 [7]. Previous bioinformatics analyses have shown that NifF [45], a Fld that can donate electrons to Nif [90], [91], [70], was recruited to *nif* operons during the transition of Nif from anaerobic to aerobic metabolic backgrounds [45]. This may point to the use of Flds as an adaptive strategy to fix N_2 in oxic environments. Moreover, under iron-deficient conditions that characterize most circumneutral oxic environments, diazotrophs tend to synthesize Flds preferentially as primary electron donors to Nif [92]. Studies have also shown that electron delivery by Fd or Fld can be complemented by other Fds or Flds that are encoded in the genomes of diazotrophs [70], [93], [94], [73], [95]. These observations suggest that pathways that mediate electron flow to nitrogenase are complex and are likely constrained by the ecological conditions encountered by the cells and the physiological background of taxa.

To better define the electron transfer system to nitrogenase in physiologically diverse microbes, Poudel et al., 2018 [7] compiled all Fd and Fld homologs in Nif-encoding genomes and classified them using homology-based methods. In addition, homologs of all enzymes that have been shown to reduce Fd or Fld for use in the reduction of N_2 by nitrogenase were compiled, including PFOR, [NiFe]-hydrogenase, [FeFe]-hydrogenase, Fix, Rnf, and both forms of FNR. Statistical analyses were then applied to this curated dataset to identify patterns of co-occurrence between the distribution of nitrogenase lineages/isoforms, enzymes that putatively reduce Fd/Fld, and Fds/Flds. These correlations were used to identify putative electron transfer pathways to Nif.

The inferred electron delivery systems to Nif is inferred to have changed markedly during the diversification of diazotrophs including at the level of the primary electron donors that provide reductant to Nif. Whereas early evolving chemotrophic anaerobic diazotrophs likely supporting N_2 reduction via oxidation of H_2 or pyruvate, more recently evolved aerobic/facultatively anaerobic diazotrophs likely supported this activity via reduced Fd supplied by the coupled oxidation of NADH/NADPH and quinone via the activity of Fix or Rnf [7]. The enzymes used to reduce Fd also likely changed as Nif diversified, with chemotrophic and anaerobic taxa likely to be dependent on [NiFe]-hydrogenase and PFOR and aerobic, facultative anaerobic or anoxygenic phototrophic taxa largely dependent on Fix and to a lesser extent PFOR and Rnf. Lastly, a different suite of Fds/Flds in aerobic/facultatively anaerobic diazotrophs were observed when compared to anaerobic diazotrophic taxa. These data are consistent with the extreme sensitivity of nitrogenase to oxygen [8] and the influence that oxygen had on the evolution of nitrogenase and nitrogen fixing organisms [45]. Together, these observations suggest that the evolution of nitrogenase has been shaped by feedbacks between the availability

of O_2 and selection to integrate O_2 and/or light into the energy metabolism of host cells while maintaining or evolving suitable pathways of capturing low potential electrons for reduction of dinitrogen in a variety of metabolic backgrounds.

CONCLUSION

Fixed N is required for all forms of life on Earth and is produced today by both abiotic and biotic mechanisms. However, on early Earth, the primary source of fixed N is likely to have been abiotic and the limited flux provided by these sources likely constrained the expansion of the biosphere thereby providing a strong selective pressure for the emergence of biological nitrogen fixation. The history of nitrogen availability is not something that can, as of yet, be ascertained definitively from the geologic record so the true time frame when fixed nitrogen limitation drove the emergence of such a complicated biochemical process is unclear. Some insights, however, can be assembled from evolutionary relationships of nitrogenase with other paralogous protein complexes associated with chlorophyll and bacteriochlorophyll biosynthesis, and the related enzyme complex proposed to be involved in cofactor F_{430} biosynthesis present in methanogens.

Phylogenetic- and structure-based examinations indicates that biological nitrogen fixation was not a property of the LUCA but rather emerged in an anaerobic archaeon and later diversified into an anaerobic bacterium and ultimately was laterally transferred into aerobic bacteria. There has been historical interest in which form of contemporary nitrogenase (Nif, Vnf, and Anf) was first to emerge. It had been proposed previously that the alternative forms of nitrogenase (Vnf and Anf) as they are termed, were likely to have evolved first due to limited available forms of Mo in the reducing early Earth. More recent phylogenetic studies, however,

indicate Nif to be the most ancestral form of these nitrogenase which likely emerged in an anaerobic, thermophilic methanogen.

Phylogenetic reconstructions of Nif proteins, in the context of proteins that are required to synthesize chlorophyll in anoxygenic and oxygenic phototrophs, allowed for a time calibration of a tree containing homologs of Anf/Vnf/Nif. These data indicate that Nif emerged ~ 2.1 Ga and was preceded by a proto nitrogenase enzyme that comprised on NifHDK (no NifEN maturases) that emerged ~ 3.5 Ga. This proto nitrogenase might have been responsible for N_2 fixation in the Archean at a time when isotopic evidence suggests biological N_2 fixation to have been present. The duplication of an ancestor of *nifDK* to yield *nifEN* at ~ 2.1 Ga allowed for refinement of the proto nitrogenase into Nif which would presumably would have been a more superior N_2 reductase. Isotopic evidence for a coupled nitrification-denitrification cycle at 2.5 Ga suggests that NH_3 was not necessarily limiting globally at this time. The advent of nitrification and denitrification, made possible by the continued oxygenation of the biosphere, would have also been a major sink on the fixed N pool and may have been the ecological trigger that led to the refinement of Nif from the proto nitrogenase.

The advent of oxygenic photosynthesis and the accompanied rise of oxygen in the Earth's atmosphere had a profound impact on the evolution of the oxygen sensitive nitrogenase enzyme. This is reflected in both phylogenetic reconstructions of NifHDK proteins, which reveal a clear demarcation between Nif in anaerobes and aerobes, as well as in composition and number of *nif* encoded genes, a set of observations that are not mutually exclusive of one another. The transition of Nif from anaerobic to aerobic metabolic backgrounds was accompanied by both gene recruitment and gene loss, resulting in an overall substantial increase in the number of *nif* genes. The majority of genes recruited during the transition from anaerobic to aerobic

nitrogen fixation are associated with metal cluster biosynthesis and regulation. Consistent with this hypothesis, the transition of Nif from anoxic to oxic environments is associated with a shift from posttranslational regulation in anaerobes to transcriptional regulation in obligate aerobes and facultative anaerobes. Thus, the mode of regulation was refined during the transition of Nif from anaerobic to aerobic metabolism likely due to selection for increased efficiency in the synthesis of Nif to meet the higher fixed N demands associated with more productive aerobic metabolism and to more efficiently regulate Nif under oxic conditions that favor protein turnover. Similarly, large changes in the pathways of electron delivery to Nif accompanied its diversification from anaerobic to aerobic backgrounds. Proteins that putatively reduce Fd with hydrogen or pyruvate were enriched in anaerobes, while those that reduce Fd with NADH/NADPH were enriched in aerobes, facultative anaerobes, and anoxygenic phototrophs. At least two mechanisms were required in aerobic, facultative anaerobic, and anoxygenic phototrophic taxa to drive reduction of Fd with higher potential NADH/NADPH, including Fix and Rnf. Given that fixed nitrogen typically limits ecosystem productivity and nitrogenase functions to relieve this limitation, the observations presented here highlight the complex interplay and feedbacks between the evolution of Earth's oxygen, nitrogen, and carbon biogeochemical cycles.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

J.W.P. thanks the National Science Foundation [grant number NSF-1331098] for supporting the work on gene expression and physiological studies. A portion of this work (structure function studies: principal investigators J. W. P. and E.S.B.) was supported as part of the Biological Electron Transfer and Catalysis Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences under Award # DE-SC0012518. E.S.B. acknowledges support from the NASA Astrobiology Institute under award # NNA15BB02A.

References

[1] R.M. Navarro-González, C. P. Nna, Mvondo D., A possible nitrogen crisis for Archaean life due to reduced nitrogen fixation by lightning., Nature, 412 (2001).
[2] J.F. Kasting, J.L. Siefert, Biogeochemistry. The nitrogen fix., Nature, 412 (2001) 26-27.
[3] J. Raymond, J.L. Siefert, C.R. Staples, R.E. Blankenship, The natural history of nitrogen fixation., Mol Biol Evol, 21 (2004) 541-554.
[4] E.S. Boyd, A.D. Anbar, S. Miller, T.L. Hamilton, M. Lavin, J.W. Peters, A late methanogen origin for molybdenum-dependent nitrogenase., Geobiology, 9 (2011) 221-232.
[5] C. Masson-Boivin, E. Giraud, X. Perret, J. Batut, Establishing nitrogen-fixing symbiosis with legumes: how many *rhizobium* recipes?, Trends Microbiol, 17 (2009) 458-466.
[6] E.S. Boyd, T.L. Hamilton, J.W. Peters, An alternative path for the evolution of biological nitrogen fixation., Front Microbiol, 2 (2011) 205.
[7] S. Poudel, E.C. Dunham, M.R. Lindsay, M.J. Amenabar, E.M. Fones, D.R. Colman, E.S. Boyd, Origin and Evolution of Flavin-Based Electron Bifurcating Enzymes., Front Microbiol, 9 (2018) 1762.
[8] L.M. Rubio, P.W. Ludden, Biosynthesis of the iron-molybdenum cofactor of nitrogenase., Annu Rev Microbiol, 62 (2008) 93-111.
[9] D.A. Betancourt, T.M. Loveless, J.W. Brown, P.E. Bishop, Characterization of diazotrophs containing Mo-independent nitrogenases, isolated from diverse natural environments., Appl Environ Microbiol. , 74 (2008) 3471-3480.
[10] A.D. Scott, V. Pelmenschikov, Y. Guo, L. Yan, H. Wang, J.G. Simon, C.H. Dapper, W.E. Newton, Y.Y. Yoda, Y. Tanaka, S.P. Cramer, Structural Characterization of CO-Inhibited Mo-Nitrogenase by Combined Application of Nuclear Resonance Vibrational Spectroscopy, Extended X ray Absorption Fine Structure, and Density Functional Theory: New Insights into the Effects of CO Binding and the Role of the Interstitial Atom., J. Am. Chem. Soc. , 136 (2014) 15942−15954.
[11] P.S. Kessler, J. McLarnan, J.A. Leigh, Nitrogenous phylogeny and the molybdenum dependence of nitrogen fixation in *Methanococcus maripaludis*., J. Bacteriol. , 179 (1997) 541–543.
[12] T.L. Hamilton, M. Ludwig, R. Dixon, E.S. Boyd, P.C. Dos Santos, J.C. Setubal, D.A. Bryant, D.R. Dean, J.W. Peters, Transcriptional profiling of nitrogen fixation in *Azotobacter vinelandii*., J Bacteriol, 193 (2011) 4477-4486.

[13] I. Severin, V. Confurius-Guns, L.J. Stal, Effect of salinity on nitrogenase activity and composition of the active diazotrophic community in intertidal microbial mats., Arch Microbiol, 194 (2012) 483-491.
[14] J. Zhu, W.J. Brill, Temperature sensitivity of the regulation of nitrogenase synthesis by *Klebsiella pneumoniae.*, J Bacteriol, 145 (1981) 1116-1118.
[15] X. Zhang, D.L. McRose, R. Darnajoux, J.P. Bellenger, F.M.M. Morel, A.M.L. Kraepie, Alternative nitrogenase activity in the environment and nitrogen cycle implications., Biogeochemistry, 127 (2016) 189–198.
[16] R. Fani, R. Gallo, P. Liò, Molecular evolution of nitrogen fixation: the evolutionary history of the *nifD*, *nifK*, *nifE*, and *nifN* genes., J. Mol. Evol. , 51 (2000) 1–11.
[17] H.C. Urey, On the Early Chemical History of the Earth and the Origin of Life., Proc Natl Acad Sci U S A., 38 (1952) 351-363.
[18] Y.L. Yung, M.B. McElroy, Fixation of nitrogen in the prebiotic atmosphere., Science, 203 (1979) 1002-1004.
[19] M.A. Schoonen, Y. Xu, Nitrogen reduction under hydrothermal vent conditions: implications for the prebiotic synthesis of C-H-O-N compounds., Astrobiology, 1 (2001) 133-142.
[20] E.S. Boyd, J.W. Peters, New insights into the evolutionary history of biological nitrogen fixation., Front Microbiol, 4 (2013) 201.
[21] E.L.B.E.S. Shock, Principles of geobiochemistry., Elements, 11 (2015) 395-401.
[22] M. Dorr, J. Kassbohrer, R. Grunert, G. Kreisel, W.A. Brand, R.A. Werner, H. Geilmann, C. Apfel, C. Robl, W. Weigand, A possible prebiotic formation of ammonia from dinitrogen on iron sulfide surfaces., Angew Chem Int Ed Engl, 42 (2003) 1540-1543.
[23] C.R. Staples, S. Lahiri, J. Raymond, L. Von Herbulis, B. Mukhophadhyay, R.E. Blankenship, Expression and Association of Group IV Nitrogenase NifD and NifH Homologs in the Non-Nitrogen-Fixing Archaeon *Methanocaldococcus jannaschii*., JOURNAL OF BACTERIOLOGY, 189 (2007) 7392–7398.
[24] K. Zheng, P.D. Ngo, V.L. Owens, X.P. Yang, S.O. Mansoorabadi, The biosynthetic pathway of coenzyme F430 in methanogenic and methanotrophic archaea., Science, 354 (2016) 339-342.
[25] S.J. Moore, S.T. Sowa, C. Schuchardt, E. Deery, A.D. Lawrence, J.V. Ramos, S. Billig, C. Birkemeyer, P.T. Chivers, M.J. Howard, S.E. Rigby, G. Layer, M.J. Warren, Elucidation of the biosynthesis of the methane catalyst coenzyme F430., Nature, 543 (2017) 78-82.
[26] S.J. Moore, S.T. Sowa, C. Schuchardt, E. Deery, A.D. Lawrence, J.V. Ramos, S. Billig, C. Birkemeyer, P.T. Chivers, M.J. Howard, S.E.J. Rigby, G. Layer, M.J. Warren, Corrigendum: Elucidation of the biosynthesis of the methane catalyst coenzyme F430., Nature, 545 (2017) 116.
[27] V. Beaunont, F. Robert, Nitrogen isotope ratios of kerogens in Precambrian cherts: a record of the evolution of atmosphere chemistry?, Precambrian Research, 96 (1999) 63-82.
[28] L.V. Godfrey, P.G. Falkowski, The cycling and redox state of nitrogen in the Archaean ocean., Nat Geosci, 2 (2009) 725–729.
[29] J. Garvin, R. Buick, A.D. Anbar, G.L. Arnold, A.J. Kaufman, Isotopic evidence for an aerobic nitrogen cycle in the latest Archean., Science, 323 (2009) 1045-1048.
[30] E.E. Stueken, R. Buick, B.M. Guy, M.C. Koehler, Isotopic evidence for biological nitrogen fixation by molybdenum-nitrogenase from 3.2 Gyr., Nature, 520 (2015) 666-669.
[31] M.C. Weiss, F.L. Sousa, N. Mrnjavac, S. Neukirchen, M. Roettger, S. Nelson-Sathi, W.F. Martin, The physiology and habitat of the last universal common ancestor., Nat Microbiol, 1 (2016) 16116.
[32] B.B. Soboh, E.S.; Zhao, D., Peters J.W.; Rubio, L,M., Substrate specificity and evolutionary implications of a NifDK enzyme carrying NifB-co at its active site., FEBS Lett. , 16 (2010) 1487-1492.
[33] L. Curatti, J.A. Hernandez, R.Y. Igarashi, B. Soboh, D. Zhao, L.M. Rubio, *In vitro* synthesis of the iron-molybdenum cofactor of nitrogenase from iron, sulfur, molybdenum, and homocitrate using purified proteins., Proc Natl Acad Sci U S A, 104 (2007) 17626-17631.
[34] P.Z. Kozbial, A.R. Mushegian, Natural history of S-adenosylmethionine-binding proteins., BMC Struct Biol, 5 (2005) 19.

[35] H.J. Sofia, G. Chen, B.G. Hetzler, J.F. Reyes-Spindola, N.E. Miller, Radical SAM, a novel protein superfamily linking unresolved steps in familiar biosynthetic pathways with radical mechanisms: functional characterization using new analysis and information visualization methods., Nucleic Acids Res, 29 (2001) 1097-1106.
[36] B.R. Duffus, T.L. Hamilton, E.M. Shepard, E.S. Boyd, J.W. Peters, J.B. Broderick, Radical AdoMet enzymes in complex metal cluster biosynthesis., Biochim Biophys Acta, 1824 (2012) 1254-1263.
[37] H. Hsu-Kim, K.M. Mullaugh, J.J. Tsang, M. Yucel, G.W.r. Luther, Formation of Zn- and Fe-sulfides near hydrothermal vents at the Eastern Lau Spreading Center: implications for sulfide bioavailability to chemoautotrophs., Geochem Trans., 9 (2008).
[38] M. Yücel, A. Gartman, C.S. Chan, G.W. Luther, Hydrothermal vents as a kinetically stable source of iron-sulphide-bearing nanoparticles to the ocean., Nat. Geosci. , 4 (2011) 1-5.
[39] A.D. Anbar, Oceans. Elements and evolution, Science, 322 (2008) 1481-1483.
[40] A.D. Anbar, Y. Duan, T.W. Lyons, G.L. Arnold, B. Kendall, R.A. Creaser, A.J. Kaufman, G.W. Gordon, C. Scott, J. Garvin, R. Buick, A whiff of oxygen before the great oxidation event?, Science, 317 (2007) 1903-1906.
[41] R.D. Joerger, R. Premakumar, P.E. Bishop, Tn5-induced mutants of *Azotobacter vinelandii* affected in nitrogen fixation under Mo-deficient and Mo-sufficient conditions., J Bacteriol, 168 (1986) 673-682.
[42] C. Kennedy, D. Dean, The *nifU*, *nifS* and *nifV* gene products are required for activity of all three nitrogenases of *Azotobacter vinelandi*i., Mol Gen Genet, 231 (1992) 494-498.
[43] A.L. Zerkle, S. Mikhail, The geobiological nitrogen cycle: From microbes to the mantle., Geobiology, 15 (2017) 343-352.
[44] G.R. Helz, C.V. Miller, J.M. Charnock, J.F.W. Mosselmans, R.A.D. Pattrick, C.D. Garner, D.J. Vaughan, Mechanism of molybdenum removal from the sea and its concentration in black shales: EXAFS evidence., Geochimica et Cosmochimica Acta, 60 (1996) 3631–3642.
[45] E.S. Boyd, A.M. Costas, T.L. Hamilton, F. Mus, J.W. Peters, Evolution of molybdenum nitrogenase during the transition from anaerobic to aerobic metabolism., J Bacteriol, 197 (2015) 1690-1699.
[46] P.C. Dos Santos, Z. Fang, S.W. Mason, J.C. Setubal, R. Dixon, Distribution of nitrogen fixation and nitrogenase-like sequences amongst microbial genomes., BMC Genomics, 13 (2012) 162.
[47] B. Dridi, M.L. Fardeau, B. Ollivier, D. Raoult, M. Drancourt, *Methanomassiliicoccus luminyensis* gen. nov., sp. nov., a methanogenic archaeon isolated from human faeces., Int J Syst Evol Microbiol, 62 (2012) 1902-1907.
[48] G.R. Helz, E. Bura-Nakićb, N. Mikacb, I. Ciglenečk, New model for molybdenum behavior in euxinic waters., Chemical Geology, 284 (2011) 323-332.
[49] B.E. Erickson, G.R. Helz, Molybdenum(VI) speciation in sulfidic waters:. Stability and lability of thiomolybdates., Geochimica et Cosmochimica Acta, 64 (2000) 1149-1158.
[50] G.R. Helz, B.E. Erickson, T.P. Vorlicek, Stabilities of thiomolybdate complexes of iron; implications for retention of essential trace elements (Fe, Cu, Mo) in sulfidic waters., Metallomics, 6 (2014) 1131-1140.
[51] G.R. Helz, T.P. Vorlicek, M.D. Kahn, Molybdenum scavenging by iron monosulfide., Environ Sci Technol, 38 (2004) 4263-4268.
[52] R.W. Kelley, J.R. Reed, W.L. Backes, Effects of ionic strength on the functional interactions between CYP2B4 and CYP1A2., Biochemistry, 44 (2005) 2632-2641.
[53] R. Price, E.S. Boyd, T.M. Hoehler, L.M. Wehrmann, E. Bogason, H.Þ. Valtýsson, J. Örlygsson, B. Gautason, J.P. Amend, Alkaline vents and steep Na^+ gradients from ridge-flank basalts—Implications for the origin and evolution of life: Geology., Ge-ology, 45 (2017).
[54] T.W. Lyons, C.T. C. T. Reinhard, N.J. Planavsky, The rise of oxygen in Earth's early ocean and atmosphere. , Nature, 506 (2014) 307–315.
[55] J.R. Gallon, The oxygen sensitivity of nitrogenase: a problem for biochemists and micro-organisms., Trends in Biochemical Sciences, 6 (1981) 19-23.
[56] J.R. Gallon, Reconciling the incompatible: N2 fixation And O2., New Physiologist, 122 (1992) 571-609.

[57] B. Bergman, J.R. Gallon, A.N. Rai, L.J. Stal, N2 Fixation by non-heterocystous cyanobacteria., FEMS Microbiology Reviews, 19 (1997) 139-185.
[58] J.C. Setubal, P. dos Santos, B.S. Goldman, H. Ertesvåg, G. Espin, L.M. Rubio, S. Valla, N.F. Almeida, D. Balasubramanian, L. Cromes, L. Curatti, Z. Du, E. Godsy, B. Goodner, K. Hellner-Burris, J.A. Hernandez, K. Houmiel, J. Imperial, C. Kennedy, T.J. Larson, P. Latreille, L.S. Ligon, J. Lu, M. Maerk, N.M. Miller, S. Norton, I.P. O'Carroll, I. Paulsen, E.C. Raulfs, R. Roemer, J. Rosser, D. Segura, S. Slater, S.L. Stricklin, D.J. Studholme, J. Sun, C.J. Viana, E. Wallin, B. Wang, C. Wheeler, H. Zhu, D.R. Dean, R. Dixon, …, D. Wood, Journal of bacteriology, 191 (2009) 4534-4545.
[59] A.S. Steunou, D. Bhaya, M.M. Bateson, M.C. Melendrez, D.M. Ward, E. Brecht, J.W. Peters, M. Kühl, A.R. Grossman, *In situ* analysis of nitrogen fixation and metabolic switching in unicellular thermophilic cyanobacteria inhabiting hot spring microbial mats., PNAS, 103 (2006) 2398-2403.
[60] P. Fay, Oxygen relations of nitrogen fixation in cyanobacteria. , Microbiol Rev., 56 (1992) 340-373.
[61] J.A. Downie, Legume nodulation., Current Biology, 24 (2014) 184-190.
[62] T. Ott, J.T. van Dongen, C. Günther, L. Krusell, G. Desbrosses, H. Vigeolas, V. Bock, T. Czechowski, P. Geigenberger, M.K. Udvardi, Symbiotic leghemoglobins are crucial for nitrogen fixation in legume root nodules but not for general plant growth and development., Current Biology, 15 (2005) 531-535.
[63] L. Soto-Urzúa, B.E. Baca, Mechanisms for protecting nitrogenase from inactivation by oxygen., Rev Latinoam Microbiol., 43 (2001) 37-49.
[64] R.K. Poole, S. Hill, Respiratory protection of nitrogenase activity in *Azotobacter vinelandii* - roles of the terminal oxidases., Biosci Rep, 17 (1997) 303-317.
[65] J. Oelze, Respiratory protection of nitrogenase in *Azotobacter* species: is a widely held hypothesis unequivocally supported by experimental evidence?, FEMS Microbiol Rev, 24 (2000) 321-333.
[66] W. Sabra, A.P. Zeng, H. Lunsdorf, W.D. Deckwer, Effect of oxygen on formation and structure of *Azotobacter vinelandii* alginate and its role in protecting nitrogenase., Appl Environ Microbiol, 66 (2000) 4037-4044.
[67] T. Egener, D.E. Martin, A. Sarkar, B. Reinhold-Hurek, Role of a ferredoxin gene cotranscribed with the nifHDK operon in N_2 fixation and nitrogenase "switch-off" of *Azoarcus* sp. Strain BH72., J Bacteriol, 183 (2001) 3752-3760.
[68] G. Sandmann, M.L. Peleato, M.F. Fillat, M.C. Lazaro, C. Gomezmoreno, Consequences of the iron-dependent formation of ferredoxin and flavodoxin on photosynthesis and nitrogen fixation on *Anabaena* Strains., Photosynth Res 26 (1990) 119-125.
[69] A.F. Yakunin, P.C. Hallenbeck, Purification and characterization of pyruvate oxidoreductase from the photosynthetic bacterium *Rhodobacter capsulatus*., Biochimica Et Biophysica Acta-Bioenergetics 1409 (1998) 39-49.
[70] A.E. Martin, B.K. Burgess, S.E. Iismaa, C.T. Smartt, M.R. Jacobson, D.R. Dean, Construction and characterization of an *Azotobacter vinelandii* strain with mutations in the genes encoding flavodoxin and ferredoxin I., (1989).
[71] Z.Y. Yang, R. Ledbetter, S. Shaw , N. Pence, M. Tokmina-Lukaszewska, B. Eilers, Q. Guo, N. Pokhrel, V.L. Cash, D.R. Dean, Evidence that the Pi release event is the rate- limiting step in the nitrogenase catalytic cycle. , Biochemistry, 55 (2016) 3625-3635.
[72] G. Gennaro, P. Hubner, U. Sandmeier , A.F. Yakunin, P.C. Hallenbeck, Cloning, characterization, and regulation of nifF from *Rhodobacter capsulatus*. , J Bacteriol, 178 (1996) 3949-3952.
[73] K. Saeki, Electron transport to nitrogenase: Diverse routes for a common destination., Springer, Place Published, 2004.
[74] D. Nieva-Gomez, G.P. Roberts, S. Klevickis, W.J. Brill, Electron transport to nitrogenase in *Klebsiella pneumoniae*., Proc Natl Acad Sci U S A, 77 (1980) 2555-2558.
[75] R.C. Wahl, W.H. Orme-Johnson, Clostridial pyruvate oxidoreductase and the pyruvate-oxidizing enzyme specific to nitrogen fixation in *Klebsiella pneumoniae* are similar enzymes. , J Biol Chem 262 (1987) 10489-10496.

[76] J.B. Therien, J.H. Artz, S. Poudel, T.L. Hamilton, Z. Liu, S.M. Noone, M.W.W. Adams, P.W. King, D.A. Bryant, E.S. Boyd, J.W. Peters, The Physiological Functions and Structural Determinants of Catalytic Bias in the [FeFe]-Hydrogenases CpI and CpII of *Clostridium pasteurianum* Strain W5., Front Microbiol. , 8 (2017) 1-11.
[77] H. Bothe, G. Neuer , I. Kalbe, G. Eisbrenner, Nitrogen Fixation., Place Published, 1980.
[78] C. Greening, A. Biswas, C.R. Carere, C.J. Jackson, M.C. Taylor, M.B. Stott, G.M. Cook, S.E. Morales, Genomic and metagenomic surveys of hydrogenase distribution indicate H_2 is a widely utilised energy source for microbial growth and survival., ISME J, 10 (2016).
[79] K. Gutekunst, X. Chen, K. Schreiber, U. Kaspar, S. Makam, J. Appel, The bidirectional NiFe-hydrogenase in *Synechocystis sp.* PCC 6803 is reduced by flavodoxin and ferredoxin and is essential under mixotrophic, nitrate-limiting conditions., J Biol Chem 289 (2014) 1930-1937.
[80] N. Khanna, P. Lindblad, Cyanobacterial hydrogenases and hydrogen metabolism revisited: Recent progress and future prospects., Int J Mol Sci, 16 (2015) 10537-10561.
[81] O. Schmitz, G. Boison, R. Hilscher, B. Hundeshagen, W. Zimmer, F. Lottspeich, H. Bothe, Molecular biological analysis of a bidirectional hydrogenase from cyanobacteria., Eur J Biochem, 233 (1995) 266-276.
[82] F.E. Rey, Y. Oda, C.S. Harwood, Regulation of uptake hydrogenase and effects of hydrogen utilization on gene expression in *Rhodopseudomonas palustris*., (2006).
[83] P.L. Dutton, Energy Transduction in Anoxygenic Photosynthesis., Photosynthesis III, 19 (1986) 197-237.
[84] B. Jagannathan, J.H. Golbeck, Photosynthesis: Microbial., Elsevier, Place Published, 2009.
[85] G. Herrmann, E. Jayamani, G. Mai, W. Buckel, Energy conservation via electron-transferring flavoprotein in anaerobic bacteria, J Bacteriol, 190 (2008) 784-791.
[86] E. Biegel, S. Schmidt, J.M. González, V. Müller, Biochemistry, evolution and physiological function of the Rnf complex, a novel ion-motive electron transport complex in prokaryotes. , Cell Mol Life Sci, 68 (2011) 613-634.
[87] B. Schrautemeier, H. Böhme, A distinct ferredoxin for nitrogen fixation isolated from heterocysts of the cyanobacterium *Anabaena variabilis*., FEBS Lett 184 (1985) 304-308.
[88] S. Apte, P. Rowell, W. Stewart, Electron donation to ferredoxin in heterocysts of the N2-fixing alga *Anabaena cylindrica*., Proc Royal Soc B, 200 (1978) 1-25.
[89] A. Aliverti, R. Faber, C.M. Finnerty, C. Ferioli, V. Pandini, A. Negri, P.A. Karplus, G. Zanetti, Biochemical and Crystallographic Characterization of Ferredoxin-NADP(+) Reductase from Nonphotosynthetic Tissues., Biochemistry, 40 (2001) 14501-14508.
[90] L.T. Bennett, M.R. Jacobson, D.R. Dean, Isolation, sequencing, and mutagenesis of the nifF gene encoding flavodoxin from *Azotobacter vinelandii*., J Biol Chem 263 (1988) 1364-1369.
[91] S. Hill, How is nitrogenase regulated by oxygen? , FEMS Microbiology Reviews, 34 (1988) 111-130.
[92] K.R. Fixen, N. Pal Chowdhury, M. Martinez-Perez, S. Poudel, E.S. Boyd, C.S. Harwood, The path of electron transfer to nitrogenase in a phototrophic alpha-proteobacterium, Environ Microbiol, 20 (2018) 2500-2508.
[93] D.C. Yoch, D.I. Arnon, Two biologically active ferredoxins from the aerobic nitrogen-fixing bacterium, *Azotobacter vinelandii*., J Biol Chem 247 (1972) 4514-4520.
[94] Y. Jouanneau, C. Meyer, I. Naud, W. Klipp, Characterization of an fdxN mutant of *Rhodobacter capsulatus* indicates that Ferredoxin-I serves as electron donor to nitrogenase., Biochim Biophys Acta-Bioenerg 1232 (1995) 33-42.
[95] V.K. Shah , G. Stacey, W.J. Brill, Electron transport to nitrogenase - Purification and characterization of pyruvate, flavodoxin oxidoreductase, the *nifJ*-gene product., J Biol Chem, 258 (1983) 2064-2068.

Figures

Figure 1. Schematic of the reduction of N_2 to NH_3 by Nif, with the reaction stoichiometry indicated. (A) Electrons carried by flavodoxin (or ferredoxin; yellow) are transferred to NifH (brown), which interacts with and transfers electrons to the P cluster, located between NifD (green) and NifK (blue), and ultimately to the FeMo-co cluster of NifD, where N_2 reduction occurs. (B) Structure and composition of *nif* gene clusters in *Azotobacter vinelandii* AVoP (both major and minor *nif* clusters) and *Methanocaldococcus* sp. FS406-22. Asterisks and boldfaced letters delineate the minimal complement of genes required to form an active nitrogenase (*nifHDKEB*). Figure adapted from Boyd et al., 2015.

Figure 2. Schematic illustrating the relationships between paralogous proteins involved in F_{430} biosynthesis (NflD), nitrogen fixation (Nif/Vnf/AnfD) and (bacterio)chlorophyll biosynthesis (Bch/ChlN). The putative co-factors and substrates for each enzyme lineage is indicated. Figure adapted from Boyd et al., 2011.

Figure 3. Proposed duplication of an ancestor of *nifD* that ultimately became the ancestor of *nifK.* Ancestors of *nifD* and *nifK* then underwent an in tandem duplication to yield *nifE* and *nifN*, respectively (A). Hydrogenotrophic methanogens encode NifD and NifK proteins that branch closest to the inferred root of the NifD and NifK phylogeny (denoted by asterisk), suggesting that the duplication of the ancestor of *nifD* to yield the ancestor of *nifK* occurred in an ancestor of these organisms. Hydrogenotrophic methanogens also encode the earliest evolving NifE and NifN proteins, suggesting the duplication of ancestral *nifDK* to *nifEN* took place in an ancestor of these organisms (not shown). Figure adapted from Boyd et al., 2011.

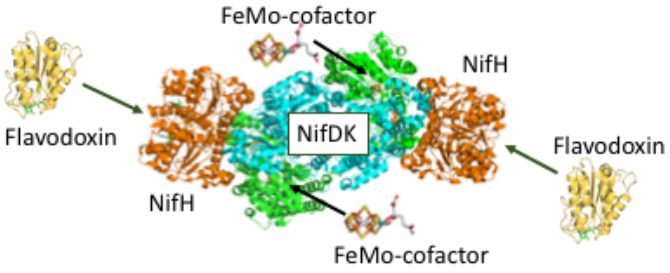
Figure 4. Schematic illustrating the possible evolution of Nif in the context of the timing of other key events. Duplication of the gene encoding the ancestor of NflD, which is thought to form a homotetramer ($_{Anc}NflD_4$) and resultant diversification resulted in the ancestor of NifD, which also is inferred to form a homotetramer ($_{Anc}NifD_4$). Duplication of the gene encoding the ancestral NifD at ~ 3.5 Ga and diversification resulted in the ancestor of NifK. The ancestral NifDK enzyme (a proto nitrogenase) likely was a heterotetramer ($_{Anc}NifD_2K_2$) and may have bound an active site co-factor similar in structure and composition to NifB-co. Duplication of the genes encoding the ancestral NifDK followed by diversification yielded the ancestor of NifEN at 2.1 Ga and allowed for the maturation of NifB-co to FeMo-co. Figure adapted from Boyd et al., 2011.

Figure 5. Two proposed models for the evolution of nitrogenase, with one specifying an origin as Nif with diversification resulting in Anf and Vnf (**A**) and the other specifying an origin as Anf with diversification resulting in Vnf and Nif (**B**). A previous phylogenetic analysis of a concatenation of Anf/Vnf/NifHDK proteins in available genome sequences, when rooted with paralogs involved in bacteriochlorophyll biosynthesis, provides evidence in support of the model depicted in panel **A** (Boyd et al., 2011).

Figure 6. Maximum Likelihood phylogenetic reconstructions from genomes are shown for (A) bacterial phylum-level groups and for (B) archaeal order- (or higher level) groups. Phylogenies were constructed using up to 104 or 31 phylogenetic marker protein coding genes, respectively (≥ 50% in any genome) from representatives of each major lineage present in the IMG/M genomic database, and the phylogenies are presented as unrooted visualizations. Groups are collapsed as triangles with the taxonomic designations provided adjacent to them. Lineages without colocalized nitrogenase protein coding genes (NifHDK) are denoted in grey, while those previously identified as containing genomes with NifHDK homologs are denoted in red, and those newly identified here are denoted in blue. The scale bar shows the expected number of substitutions per site.

Figure 7. Schematic illustrating the diversification of Nif from strictly anaerobic organisms (green lineages) to aerobes and facultative anaerobes (red lineages) and the *nif*-associated genes that exhibit the strongest evidence for gene loss (indicated by a "-") or gain (indicated by a "+") as revealed by a previous coupled bioinformatics-evolutionary analysis (A) (Boyd et al., 2015). Unique and shared genes in the nif gene clusters of aerobes/facultative anaerobes (red) and anaerobes (green); shared genes are colored purple (B). Generalized mechanisms of nitrogenase regulation in aerobes/facultative anaerobes (top) and anaerobes (bottom) (C). Low levels of ammonia, oxygen, or elevated 2 oxoglutarate (2-OG) promote interaction of NifA with the promotor or promote NifL to interact with NifA, which allows it to then interact with the promoter, thereby allowing for transcription of nitrogen fixation in aerobes/facultative anaerobes (top) (see *J Bacteriol. 2004 Feb; 186(3): 601–610* for additional details on transcriptional regulation in these taxa). In contrast, the generalized mechanism of nitrogenase regulation in anaerobes is through post-translational modification. Here, high levels of 2-OG prevent interaction of $NifI_1I_2$ with NifDK, allowing for NifH to associate with NifDK and deliver electrons for N_2 reduction activity. Low levels of 2-OG allow for $NifI_1I_2$ to associated with NifDK, preventing access of NifH to NifDK and thereby down regulating N_2 fixation (see *Proc Natl Acad Sci U S A. 2006 Jun 27; 103(26): 9779–9784* for additional details on nitrogenase regulation in anaerobes).

A.



$$N_2 + 8e^- + 8H^+ + 16MgATP \longrightarrow 2NH_3 + H_2 + 16MgATP$$

B.