Geobiology

Geobiology (2011), 9, 221–232 DOI: 10.1111/j.1472-4669.2011.00278.x

A late methanogen origin for molybdenum-dependent nitrogenase

E. S. BOYD,[1] A. D. ANBAR,[2] S. MILLER,[3] T. L. HAMILTON,[1] M. LAVIN[4] AND J. W. PETERS[1]

[1]*Department of Chemistry and Biochemistry and the Astrobiology Biogeocatalysis Research Center, Montana State University, Bozeman, MT, USA*
[2]*School of Earth and Space Exploration and Department of Chemistry and Biochemistry, Arizona State University, Tempe, AZ, USA*
[3]*Department of Biological Sciences, University of Montana, Missoula, MT, USA*
[4]*Department of Plant Sciences, Montana State University, Bozeman, MT, USA*

ABSTRACT

Mounting evidence indicates the presence of a near complete biological nitrogen cycle in redox-stratified oceans during the late Archean to early Proterozoic (*c.* 2.5–2.0 Ga). It has been suggested that the iron (Fe)- or vanadium (V)-dependent nitrogenase rather than molybdenum (Mo)-dependent form was responsible for dinitrogen fixation during this time because oceans were depleted in Mo and rich in Fe. We evaluated this hypothesis by examining the phylogenetic relationships of proteins that are required for the biosynthesis of the active site cofactor of Mo-nitrogenase in relation to structural proteins required for Fe-, V- and Mo-nitrogenase. The results are highly suggestive that among extant nitrogen-fixing organisms for which genomic information exists, Mo-nitrogenase is unlikely to have been associated with the Last Universal Common Ancestor. Rather, the origin of Mo-nitrogenase can be traced to an ancestor of the anaerobic and hydrogenotrophic methanogens with acquisition in the bacterial domain via lateral gene transfer involving an anaerobic member of the *Firmicutes*. A comparison of substitution rates estimated for proteins required for the biosynthesis of the nitrogenase active site cofactor and for a set of paralogous proteins required for the biosynthesis of bacteriochlorophyll suggests that Nif emerged from a nitrogenase-like ancestor approximately 1.5–2.2 Ga. An origin and ensuing proliferation of Mo-nitrogenase under anoxic conditions would likely have occurred in an environment where anaerobic methanogens and *Firmicutes* coexisted and where Mo was at least episodically available, such as in a redox-stratified Proterozoic ocean basin.

Received 21 October 2010; accepted 15 March 2011

Corresponding author: John W. Peters. Tel.: +1 406 994 7212; fax: +1 406 994 5407; e-mail: john.peters@chemistry.montana.edu

INTRODUCTION

Biologically available (fixed) nitrogen is essential for all of life and was probably generated on early Earth by abiotic processes such as lightning discharge (Yung & Mcelroy, 1979; Kasting & Walker, 1981). Abiotic sources of fixed nitrogen are thought to have become limiting to an expanding global biome, which may have precipitated the innovation of biological dinitrogen (N_2) fixation. Using simulations of Archaean atmospheric chemistry, Navarro-Gonzalez contend that decreasing CO_2 concentrations and concomitant decreases in abiotic N_2 reduction to nitrous oxide led to a nitrogen crisis at approximately 2.2 Ga (Navarro-González *et al.*, 2001). However, using the same logic, Kasting & Siefert (2001) argued that the nitrogen crisis could have ensued as early as 3.5 Ga.

Biological nitrogen fixation, or the reduction of N_2 to ammonia, functions to relieve fixed N limitation (Zehr *et al.*, 2003; Rubio & Ludden, 2008). This process is catalyzed in a small number of diverse bacteria and methanogenic archaea by one of three evolutionarily related nitrogenases that contain active site cofactors containing iron and molybdenum, iron and vanadium, or iron alone (Nif, Vnf, Anf respectively) (Joerger & Bishop, 1988). Evidence for an early emergence of biological nitrogen fixation exists in the isotopic record of organic-rich cherts and shales dated to >2.5 Ga which reveal $^{15}N/^{14}N$ ratios that are suggestive of biological nitrogen fixation (Beaumont & Robert, 1999; Garvin *et al.*, 2009; Godfrey & Falkowski, 2009). Prior to the rise of O_2 approximately 2.5 Ga, it has been suggested that Anf or Vnf rather than Nif were responsible for nitrogen fixation since oceans

were depleted in Mo (Anbar *et al.*, 2007; Scott *et al.*, 2008) due to the insolubility of Mo-sulphides under anoxic conditions (Helz *et al.*, 1996). However, it is unclear from past investigations of the evolutionary record which metalloenzyme was most likely responsible for N_2 fixation at this time (Raymond *et al.*, 2004; Soboh *et al.*, 2010).

Based on phylogenetic examination of nitrogenase protein sequences, two equally parsimonious origins for Mo-nitrogenase have recently been proposed (Leigh, 2000; Raymond *et al.*, 2004). One proposal invokes an origin for Mo-nitrogenase in the Last Universal Common Ancestor (LUCA origin model) (Leigh, 2000), whereas the other proposal invokes an origin for Mo-nitrogenase in the methanogenic archaea (methanogen origin model) with subsequent acquisition in the bacterial domain via lateral gene transfer (LGT) (Raymond *et al.*, 2004). Importantly, the anoxic nature of oceans prior to the rise of O_2 (LUCA origin model) and the anaerobic nature of methanogens (methanogen origin model) both require that Nif evolved in an anoxic environment. This follows from the likelihood that O_2 was scarce until approximately 2.5 Ga (Canfield *et al.*, 2000; Holland, 2002; Anbar *et al.*, 2007), which postdates when LUCA is thought to have emerged and then split to form the bacteria and archaea (Mojzsis *et al.*, 1996; Parsons *et al.*, 1998; Penny & Poole, 1999). Likewise, all methanogens examined to date are obligate anaerobes (Boone *et al.*, 1993; Thauer *et al.*, 2008), suggesting that the emergence of any enzyme at any time during the evolution of this lineage would have occurred under anoxic conditions. Importantly, both of the aforementioned scenarios which invoke an anoxic origin for Mo-nitrogenase are consistent with the extreme oxygen sensitivity of this enzyme (Rubio & Ludden, 2008).

The majority of present-day biological N_2 reduction is catalyzed by the Mo-dependent nitrogenase (encoded by *nif*) (Rubio & Ludden, 2008), although V- and Fe-nitrogenase (encoded by *vnf* and *anf* respectively) are important biological sources of fixed nitrogen in environments where Mo is limiting (Joerger & Bishop, 1988). To date, the genomes of organisms that encode for Fe- and V-nitrogenase universally encode for Mo-nitrogenase (Raymond *et al.*, 2004; Soboh *et al.*, 2010), suggesting that they function secondarily to Mo-nitrogenase in extant organisms. The Mo-dependent nitrogenase (Nif) complex consists of a homodimeric Fe protein (NifH) component that donates electrons to the heterotetrameric MoFe protein (NifDK) component, which contains the iron–molybdenum cofactor (FeMo-co) substrate reduction sites (Bulen & Lecomte, 1966). In addition to encoding for nitrogenase structural proteins, *nif* regulons encode for proteins involved in regulation and metal cofactor (e.g. FeMo-co) biosynthesis (Rubio & Ludden, 2008). Several of the genes required to synthesize FeMo-co evolved from duplication and fusion events. For example, the heterotetrameric NifEN complex functions as a required scaffold for FeMo-cofactor assembly (Ugalde *et al.*, 1984; Jacobson *et al.*, 1989; Roll *et al.*, 1995; Hu *et al.*, 2005) and is currently thought to have arisen through a paralogous duplication of *nifDK* (Fig. 1A) (Brigle *et al.*, 1987; Fani *et al.*, 2000; Raymond *et al.*, 2004; Soboh *et al.*, 2010). NifB, another protein required for FeMo-co biosyn-

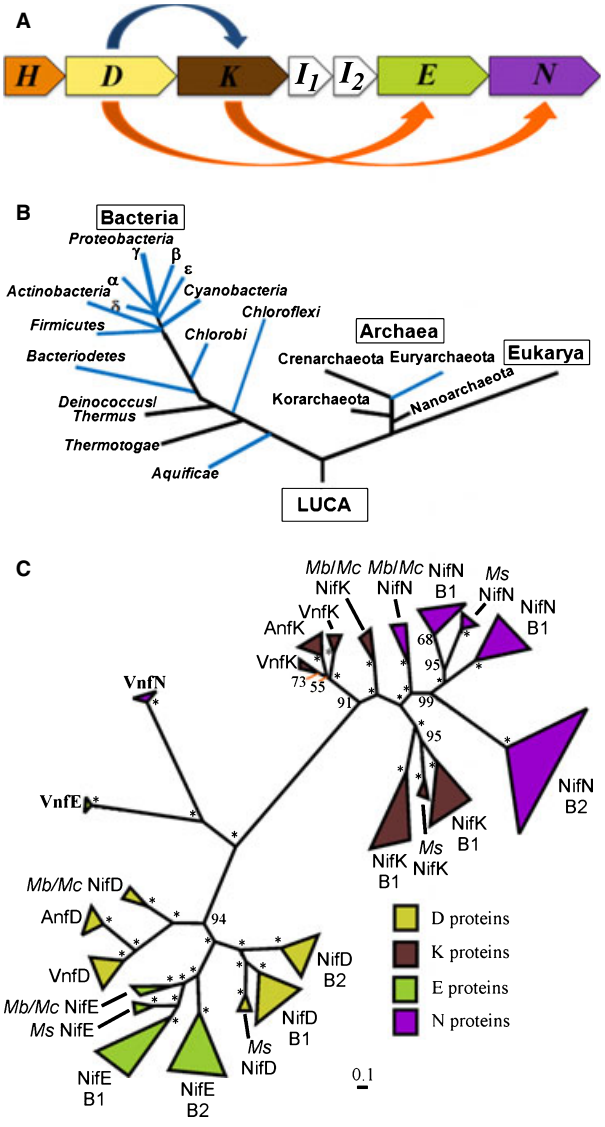

Fig. 1 (A) Nitrogenase regulon from *Klebsiella pneumoniae* with the blue arrow indicating the duplication event of *nifD* yielding *nifK* and the orange arrows indicating the in-tandem duplication event of *nifDK* yielding *nifEN*. (B) Universal tree indicating lineages where full complements of nitrogenase proteins (HDKENB) were found (blue lineages). (C) Bayesian consensus phylogenetic tree of Nif/Vnf/AnfD, K, E and N proteins from 40 taxa that represent the primary lineages (C). The width and depth of collapsed clades proportionally reflect the number and the diversity of sequences in that clade respectively. *Methanobacteriales/Methanococcales* proteins are abbreviated *Mb/Mc* and *Methanosarcinales* proteins are abbreviated by *Ms*. All lineages labelled as Nif without an *Mb/Mc* or *Ms* designation contain only bacterial sequences. Asterisks indicate 100% posterior probability support; bar equals 1 substitution per 10 sequence positions.

thesis (Christiansen *et al.*, 1998; Curatti *et al.*, 2006), catalyzes the formation of a FeMo-co precursor iron–sulphur cluster, termed NifB-co (Shah *et al.*, 1994). In most organisms, NifB is a fusion protein consisting of an amino terminal domain belonging to the radical S-adenosyl methionine (SAM) family of proteins and a carboxy terminal domain belonging to the NifX/NafY family of proteins (Rubio & Ludden, 2008).

In this work, we have focused on these specific gene duplication (e.g. Anf/Vnf/NifDKEN) and fusion (e.g. NifB) events in proteins required to synthesize an active nitrogenase, to infer the evolutionary origin of Mo-nitrogenase and its evolutionary relationship to the 'alternative' metal-containing forms of nitrogenase. In addition, we examine the temporal relationships between paralogous proteins required for the biosynthesis of the active site cofactor of nitrogenase and the biosynthesis of bacteriochlorophyll [e.g. light-independent (dark-operative) protochlorophyllide oxidoreductase (encoded by *bchNB*) and chlorophyllide reductase (encoded by *bchYZ*) (Hearst *et al.*, 1985; Burke *et al.*, 1993; Fujita *et al.*, 1993)] to provide temporal insight into the emergence of Mo-nitrogenase. *bchNB* is present in the genomes of all chlorophototrophs whereas *bchYZ* is present in the genomes of all anoxygenic phototrophs (Chew & Bryant, 2007). BchNY and BchBZ are paralogs of Anf/Vnf/NifD and Anf/Vnf/NifK, respectively, and are thought to have arisen from a common ancestral reductase by a series of gene duplications (Burke *et al.*, 1993). Therefore, the careful examination of these paralogous proteins has significant potential to provide insight into the evolutionary linkages and temporal relationships between the emergence of photosynthesis relative to nitrogenase. The results are discussed in light of the emerging perspective on redox conditions in late Archaean to early Proterozoic marine environments (Scott *et al.*, 2008; Godfrey & Falkowski, 2009; Reinhard *et al.*, 2009; Kendall *et al.*, 2010; Poulton *et al.*, 2010), at a time when circumstances were prime for the evolution of the Mo-nitrogenase (Anbar & Knoll, 2002; Anbar, 2008).

MATERIALS AND METHODS

Vnf/Anf/NifDKEN and NifB phylogenetic analyses

Blastp was used to compile all D, K, E, N and B protein sequences from genomic sequences using the DOE-IMG and the NCBI Genome Blast servers in July 2010 (Table S1). Those which contained a full complement of DKENB were retained in the case of Vnf and Nif operons, or DKB in the case of Anf operons. Each individual locus was aligned using ClustalX (version 2.0.8) with the Gonnet 250 protein matrix and default gap extension and opening penalties (Larkin *et al.*, 2007). Each alignment was scrutinized and manually aligned using known catalytic residues (Kim & Rees, 1992; Peters *et al.*, 1995; Kaiser *et al.*, 2011). ProtTest (version 2.0) (Abascal *et al.*, 2005) was used to select WAG+I+G+F as the best-fit protein evolutionary model for individual Vnf/Anf/NifDKEN and NifB protein sequences and to determine the amino acid frequency (specified as 'F' in the model). The phylogeny of each protein sequence was evaluated using MrBayes (version 3.1.2) (Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck, 2003) using the WAG evolutionary model with fixed (F) amino acid frequencies and gamma-distributed rate variation with a proportion of invariable sites (I+G). Tree topologies were sampled every 500 generations over 1.0×10^6 generations (after a burnin of 1.0×10^6) at likelihood stationarity and after convergence of two separate Markov chain Monte Carlo runs (average standard deviation of split frequencies <0.05). Taxa with D, K, E and N protein sequences that represented each of the primary lineages were empirically selected (Table S2) and these sequences were subjected to evolutionary analysis as described above (see Fig. S1 legend). A composite DKEN phylogram was constructed using the WAG evolutionary model with fixed amino acid frequencies (F) and gamma-distributed rate variation with a proportion of invariable sites (I+G). Tree topologies were sampled at likelihood stationarity as described above. The unrooted phylogram with collapsed clades (Fig. 1) was projected using TreeView (version 1.6.6) (Page, 1996). The radial phylogram (Fig. 2) was projected with FigTree (version 1.2.2) (http://tree.bio.ed.ac.uk/UH). NifB protein sequences from the taxa selected for the DKEN analysis were compiled and subjected to evolutionary analysis as described above, using the WAG+I+G+F substitution model. Tree topologies were sampled at likelihood stationarity as described above, and a consensus phylogram was projected using FigTree. The orientation of *nif* regulons from *Bradyrhizobium japonicum* and *Burkholderia phymatum* (Fig. 3B) has been reversed to conserve space. The genomic context of nitrogenase regulons was determined manually using the IMG Gene Neighborhood viewer.

Composite Vnf/Anf/NifDKEN and BchNBYZ phylogenetic analyses

Blastp was used to compile all BchNBYZ protein sequences from genomic sequences using the DOE-IMG and the NCBI Genome Blast servers in November of 2009 (Table S3). To simplify the analysis, homologues of (bacterio)chlorophyll biosynthesis proteins from eukarya were not considered in this analysis since previous studies indicate eukaryote homologs to be derived from bacterial sequences (Xiong *et al.*, 2000). Sequence alignment, evolutionary model selection and phylogenetic analyses were performed as described above. Taxa representing the primary lineages for each individual BchNBYZ and Vnf/Anf/NifDKEN protein sequence were used to construct a composite Vnf/Anf/NifDKEN/BchNBYZ phylogram. The radial phylogram (Fig. S3) was projected with FigTree.

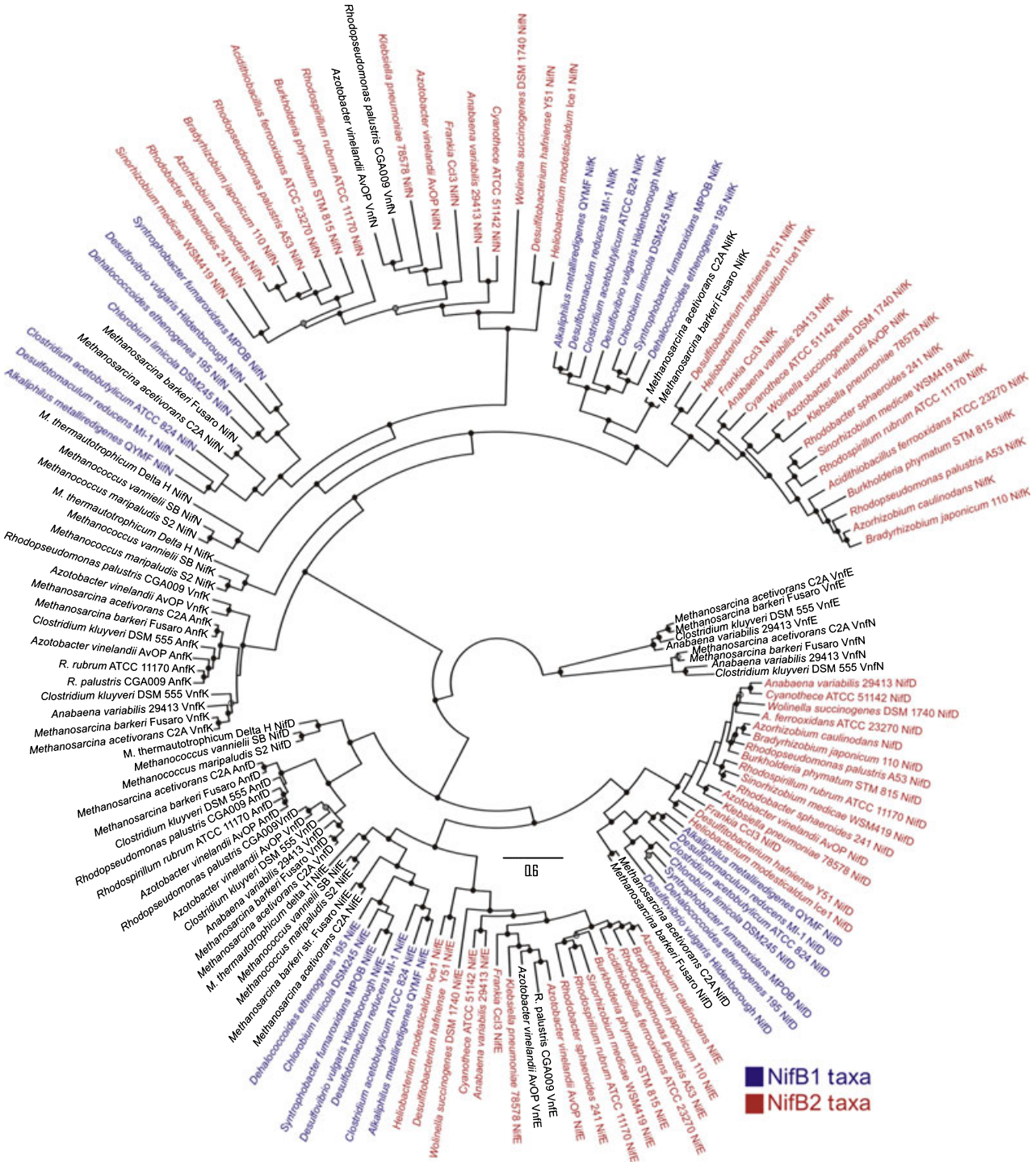

Fig. 2 Bayesian consensus phylogenetic tree of Nif/Vnf/AnfD, K, E, and N proteins from 40 taxa that represent the primary lineages for each of the four proteins. Black circles at nodes denote >90% posterior probability (PP), grey circles at nodes denote >80% PP, open circles denote >70% PP, and no symbol denotes 50–70% PP. Nodes with <50% PP were collapsed; bar equals 6 substitutions per 10 sequence positions.

Evolutionary rates analysis

Estimated rates of substitution were converted to absolute ages for individual crown clades using three independent approaches including a data-driven penalized-likelihood (PL) approach that allows for rate variation along branches of the phylogenetic tree (Sanderson, 2002), a molecular clock type approach that assumes a constant rate of evolution (Sander-

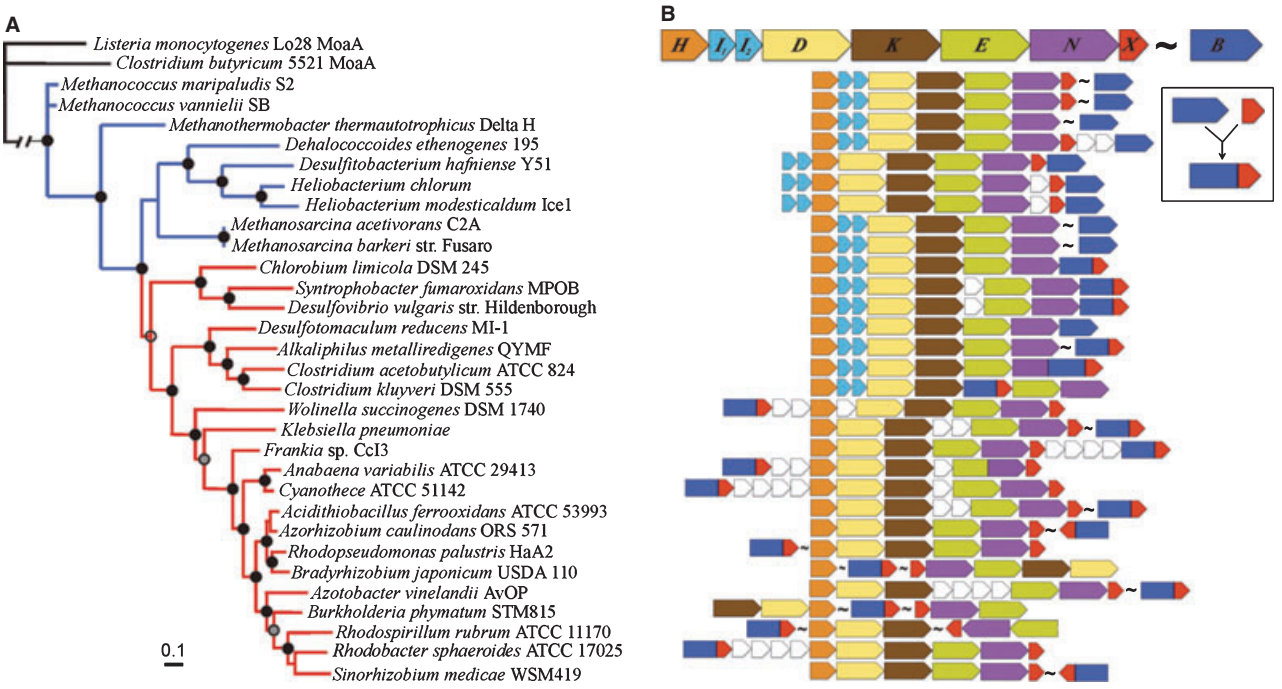

Fig. 3 (A) Phylogenetic tree of the radical SAM domain of NifB from 30 taxa that contain NifHDKEN rooted with the molybdenum biosynthesis protein (MoaA). Taxa were selected to represent the major lineages of NifB. Lineages highlighted in blue denote taxa with NifB proteins which contain only the SAM domain and lineages highlighted in red denote fused 'SAM–NifX' proteins. Black circles at nodes denote >90% posterior probability (PP), grey circles at nodes denote >80% PP, open circles denote >70% PP, and no symbol denotes 50–70% PP. Nodes with <50% PP were collapsed; bar equals 1 substitution per 10 sequence positions. (B) Genetic structure of the *nif* regulons for taxa presented in phylogenetic tree. A gene locus colour key is indicated at the top of the panel and an illustration of the NifB–NifX fusion event is inset. The orientation of *nif* regulons from *Bradyrhizobium japonicum* and *Burkholderia phymatum* has been reversed to conserve space.

son, 2004), and a nonparametric (NP) approach that seeks to minimize ancestor to descendent local rate changes (Sanderson, 1997), as implemented in version 1.71 of the program R8S (Sanderson, 2002). The 'VnfEN' lineage was used to root the tree. 'VnfEN' was chosen as it is present in both archaea and bacteria, it branches near the base of the tree. While the limitations of rate smoothing and molecular clock approaches including the non-clocklike behaviour of proteins (Pagel, 1999) and the bias towards overestimation of evolutionary time scales (Rodriguez-Trelles *et al.*, 2002) are acknowledged and appreciated, we have used a number of rate smoothing approaches (see above) as well as a range of forced time constraints to accommodate the various lines of evidence for the emergence of LUCA (root age constraint) and for the emergence of oxygenic photosynthesis (oxygenic photosynthesis age constraint). The various time constraints utilized include a fixed root age of either 3.5 or 3.8 Ga and a fixed age for the emergence of oxygenic photosynthetic lineage of 2.5 or 2.8 Ga. A root age of 3.8 Ga was chosen as isotopic evidence suggests that LUCA may have emerged around this time (Mojzsis *et al.*, 1996; Parsons *et al.*, 1998). However, evidence in support of this date for the first life has been debated, and microfossil evidence substantiating this early calibration date has yet to be identified. Thus, a second calibration point of 3.5 Ga based on the earliest documented microfossil evidence for life on Earth (Schopf, 1992) was also included in the rate estimates. The age of the nodes demarcating the divergence of oxygenic and anoxygenic phototrophic lineages in the earliest branching of the two subclusters (C1) in each of the BchN and B lineages was calibrated to a conservative age of 2.5 Ga to reflect the timing of the Great Oxidation Event (GOE) (Canfield *et al.*, 2000; Holland, 2002; Anbar *et al.*, 2007; Kaufman *et al.*, 2007; Godfrey & Falkowski, 2009; Reinhard *et al.*, 2009; Kendall *et al.*, 2010) as well as a calibration point of 2.8 Ga to accommodate additional evidence indicating that oxygenic phototrophs may have evolved by this time (Summons *et al.*, 1999; Brocks *et al.*, 2003; Eigenbrode *et al.*, 2008). Given that BchNB proteins from oxygenic phototrophs are nested among BchNB proteins derived from anoxygenic phototrophs in three of the four BchNB subclusters (Fig. S1), it is unlikely that mass extinction of oxygenic phototrophs has occurred. Thus, the crown clade ages in the earliest branching C1 subcluster of BchN and BchB reflect the minimum age estimates of the lineages and not just the extant diversification. Importantly, each rate smoothing method (PL, molecular clock, NP) yielded similar estimated age mean and variances. Since derivations of age estimates from substitution rates centre on rate inconstancy, the fact that the three independent methods produce similar results signals that the sequences are evolving closer to rate constancy. Thus, only

the results of the PL method are presented, as this is a data-driven method and the least subjective estimation of substitution rates. The PL rate-smoothed chronogram (Fig. S3) was projected with FigTree.

RESULTS AND DISCUSSION

Origin of Nif

Nitrogenase regulons containing D, K, E and N homologues were identified in the genomes of the methanogenic archaea and a variety of bacteria (Fig. 1B). Importantly, homologues of one or more of these loci are absent from many lineages of the archaea (*Crenarchaea*, *Korarchaea* and *Nanoarchaea*) and early branching members of the bacteria (*Thermotogae* and *Deinococcus/Thermus*) (Fig. 1B). Among the bacteria, the deepest branching bacterial lineage with evidence of homologues of nitrogenase occurred in *Thermocrinis albus* DSM 14484 (*Aquificae*) (Fig. 1B), although these regulons were incomplete [homologues of nifD, K and N only (Table S1)] and the branching order of individual proteins suggests that *nif* was acquired in *T. albus* by LGT with either a firmicute or a proteobacterium (data not shown). The distribution of *anf* and *vnf* operons in genome sequences was far more restricted than for *nif*. Anf operons encoding for DK were identified in the genomes of several bacterial and methanogen genomes that also encoded for Nif (Table S1), a finding that is consistent with the requirement for either *nif*-encoded E and N homologues as scaffolds in the synthesis of an active Anf nitrogenase (Wolfinger & Bishop, 1991; Rubio & Ludden, 2005, 2008). Likewise, *vnf* regulons were identified in several bacterial and methanogen genomes that also encoded for *nif* (Table S1).

The individual D, K, E and N protein sequence clusters provide a root for each other in evolutionary reconstructions (Baldauf *et al.*, 1996), as genes encoding D, K, E and N proteins resulted from gene duplication (Brigle *et al.*, 1987; Fani *et al.*, 2000; Raymond *et al.*, 2004; Soboh *et al.*, 2010). To further evaluate if Mo-nitrogenase was associated with LUCA or if it emerged following the split between bacteria and archaea, we reconstructed the evolutionary history of DKEN protein sequences and examined the branching order of archaeal and bacterial homologues for each individual sequence. The DKEN phylogram revealed distinct and well-supported lineages that generally corresponded to the metal content of the constituent proteins (Figs 1C and 2). Importantly, NifD protein sequences are paraphyletic with respect to NifE, with one lineage containing only archaeal sequences and the other containing both archaeal and bacterial sequences. Similar to NifD protein sequences, NifK sequences are paraphyletic with respect to NifN, with one lineage containing only archaeal sequences and the other with both archaeal and bacterial sequences. The NifE and NifN lineages are each monophyletic and nested within NifD and NifK lineages respectively (Figs 1C and 2). This is indicative of a duplication of a *nifDK* ancestor yielding *nifEN*, which is consistent with inferences drawn from previous phylogenetic studies of these protein sequences (Fani *et al.*, 2000; Soboh *et al.*, 2010). The phylogenetic evidence from these studies, which collective indicate that *nifDK* duplicated to *nifEN*, is counter to recent conjectures put forth based on biochemical data suggesting that NifEN might predate NifDK, or at the very least co-evolved with NifDK (Hu *et al.*, 2009, 2010). The finding that NifEN homologues are nested within Anf/Vnf/NifDK homologues provides strong support for an evolutionary scenario where the duplication of an ancestor of *nifDK* to *nifEN* enabled the further maturation of a FeMo-co precursor (e.g. NifB-co), resulting in a reductase with biochemical reactivity that was superior to the reductase synthesized in the absence of NifEN. These observations, coupled with the finding that NifEN from archaea and bacteria are not reciprocally monophyletic, provide strong support for the hypothesis that Mo-nitrogenase (as defined by the presence of *nifEN*) was not present in LUCA, and implicates a role for LGT in acquisition of *nif* in the bacterial domain (discussed below).

In individual NifD, NifK, NifE and NifN lineages, methanogens of the order *Methanococcales* (*Methanococcus maripaludis* and *Methanococcus vannielli*; denoted by *Mc*) and *Methanobacteriales* (*Methanothermobacter thermoautotrtophicus*; denoted by *Mb*) form distinct groups that diverged before all other Nif proteins, including those from the *Methanosarcinales* (*Methanosarcina barkeri* and *Methanosarcina acetivorans*; denoted by *Ms*) (Figs 1C and 2). This finding is consistent with phylogenetic analysis of other methanogen genes including *nifH* and 16S rRNA genes, where genes from members of the *Methancoccales* and *Methanobacteriales*, which utilize H_2 to reduce CO_2 to CH_4, branch earlier than the metabolically more versatile *Methanosarcinales* (Boone *et al.*, 1993; Raymond *et al.*, 2004; Thauer *et al.*, 2008). The topology of NifD and NifK lineages, where *Methanococcales/Methanobacteriales* and *Methanosarcinales/*bacterial sequences are paraphyletic with respect to NifE and NifN, suggests that the duplication of *nifDK* resulting in *nifEN* probably occurred in an ancestor of the hydrogenotrophic methanogens. Two primary lineages of bacterial NifDKEN are apparent in the phylogenetic reconstruction (denoted as NifB1 and NifB2 in Figs 1C and 2). Bacterial proteins that form NifB1 are from strictly anaerobic taxa (*Firmicutes*, *Proteobacteria*, green non-sulphur bacteria and green sulphur bacteria) and these form a sister clade to proteins from the *Methanosarcinales* in each NifDKE lineage, but interestingly are paraphyletic with respect to *Methanosarcinales* Nif-N sequences. The NifB2 cluster comprised sequences from both aerobic and anaerobic taxa, including those from the strict anaerobes *Heliobacterium modesticaldum* and *Desulfitobacterium hafniense* which consistently branch at the base of individual Nif B2 DKEN lineages (Figs 1C and 2). Together, these observations

suggest a role for LGT event(s) in the acquisition of Nif in bacteria, most likely involving an ancestral methanogen and an ancestral member of the *Firmicutes*. This interpretation is consistent with the close spatial proximity noted between members of the *Methanosarcinales* and *Firmicutes* in a variety of extant anoxic environments (Stams, 1994) and with a number of reports of LGT events involving individual genes or entire metabolic pathways between these two anaerobic lineages (Beiko *et al.*, 2005; Fournier & Gogarten, 2008). An origin for Mo-nitrogenase in strictly anaerobic and hydrogenotrophic methanogens (Boone *et al.*, 1993; Thauer *et al.*, 2008), and the early proliferation of *nif* in anaerobic bacteria is consistent with the extreme oxygen sensitivity of the nitrogenase complex (Rubio & Ludden, 2008).

To further investigate the likelihood of an archaeal origin for Mo-dependent biological nitrogen fixation, we examined the phylogeny and genetic structure of NifB from taxa that contain the full complement of NifHDKEN, since like *nifHEN*, the *nifB* gene product is essential for the biosynthesis of FeMo-co (Shah *et al.*, 1994; Christiansen *et al.*, 1998; Curatti *et al.*, 2006). Alignment of NifB sequences revealed that methanogenic archaea universally lack the C-terminal 'NifX' domain (data not shown). The phylogenetic distribution of NifB sequences containing only the SAM domain (lack the fused C-terminal 'NifX' domain) amongst bacteria was sporadic, with only a few anaerobes within the *Firmicutes*, *Chloroflexi* and *Proteobacteria* sharing this genotype. Phylogenetic reconstruction of the SAM domain of these sequences indicated that all NifB sequences lacking the NifX domain diverged very early with respect to those containing the fused SAM and NifX domains (Fig. 3A). Importantly, NifB sequences from both *Methanococcales* and *Methanobacteriales* formed well-supported and early branching lineages, indicating that they evolved early with respect to NifB from the *Methanosarcinales* and *Bacteria*. NifB from a number of *Firmicutes* formed a sister group with NifB from the *Methanosarcinales*, providing additional support for LGT in the acquisition of *nif* in the bacteria via an interaction between an ancestral member of the *Firmicutes* and an ancestral member of the *Methansarcinales*.

The genetic structure of *nif* regulons that contained *nifB* genes encoding for fused 'SAM–NifX' domains was then compared with those that contained *nifB* genes encoding for only the SAM domain. In contrast to *nif* regulons from genomes of organisms containing genes encoding for fused NifB–NifX proteins, the gene order is generally more conserved in organisms containing the early branching NifB sequences lacking the 'NifX' fusion domain (Fig. 3B). The synteny of *nif* regulons from a diversity of methanogenic genomes is consistent with the hypothesis that Mo-nitrogenase emerged in the hydrogenotrophic methanogen lineage. Moreover, the lack of the fused NifB–NifX in anaerobic members of the *Firmicutes*, as well as similar operon structures in these organisms supports the hypothesis that *nif* was acquired in the bacterial domain via LGT from a methanogen to a firmicute [for further discussion of the proliferation of *nif* genes in bacteria, we refer readers to the comprehensive works of Raymond *et al.* (2004) and more recently by Hartmann & Barnum (2010)].

Evolutionary relationship between Nif, Anf and Vnf nitrogenase

While the primary goal of this study was to investigate the origin of the Mo-dependent nitrogenase, the data also provide insight into the evolutionary relationships between the various forms of nitrogenase. The genes encoding the structural proteins (HDK) for each of the three forms of nitrogenase share high levels of sequence identity, suggesting derivation from a single ancestral nitrogenase (Kessler *et al.*, 1997; Leigh, 2000; Raymond *et al.*, 2004). NifD proteins from methanogens of the order *Methanococcales* and *Methanobacteriales* form a very early branching group that along with AnfD and VnfD, form a lineage that branches distinctly from bacterial and *Methanosarcinales* NifD. This phylogenetic relationship is nearly mirrored in the paralogous earliest K/N branching lineages where Vnf/AnfK proteins form a separate lineage that branches distinctly from the *Methanococcales* and *Methanobacteriales* NifK lineage. Evidence that VnfD diversified before AnfD is seen in the deeper crown clade of VnfD when compared to AnfD, which is mirrored in the paralogous earliest K/N branching lineage, where the well-supported AnfK clade is nested within a more diverse VnfK group. This observation may indicate that AnfDK results from an in-tandem duplication of VnfDK. This interpretation is bolstered by the low posterior probability supporting the Vnf/AnfK, which is evidence that the three sublineages could just as well form a grade at the base of the K/N lineage (Figs 1C and 2), where VnfK is resolved paraphyletic in contrast to the monophyletic AnfK.

The fact that Anf/VnfK did not form a coherent lineage with '*Mc/Mb*' NifK sequences such as that observed in the well-supported Anf/Vnf/'*Mc/Mb*' NifD lineage was surprising (Figs 1C and 2), considering that the evolutionary history of Anf/Vnf/NifD and Anf/Vnf/NifK would be expected to be congruent considering that they together form the structural component of nitrogenase (Eady, 1996; Rubio & Ludden, 2008) and would be expected to be selected for in concert. However, support for the node at the base of the Anf/VnfK lineage is low (posterior probability = 91) and empirical studies have shown that nodes with posterior probabilities of <95 can be prone to type I phylogenetic errors (Erixon *et al.*, 2003). Thus, it is possible that phylogenetic studies that include additional Anf/Vnf/NifDK sequences which become available in future genome sequencing projects will improve the phylogenetic resolution within this lineage which in turn, may reconcile the differences in branching order between the poorly supported Anf/Vnf/'*Mc/Mb*' NifK sublineages, when

compared to the well-supported Anf/Vnf/'*Mc/Mb*' NifD sublineages.

The 'VnfEN' lineage, which comprises two constituent and divergent sub-lineages, branches near the root of the DKEN phylogram (Fig. 1C), suggesting these proteins to have diverged very early with respect to Vnf/Anf/NifKEN proteins. Intriguingly, the divergent 'VnfN' protein from *Anabaena variabilis* was shown to be required for the function of Vnf nitrogenase (Thiel, 1996). Although it is not known if 'VnfN' performs an analogous scaffold role like NifN, it is important to note that NifN did not complement for the mutant *A. variabilis* strain lacking 'VnfN' (Thiel, 1996). This is in contrast to reports in *Azotobacter vinelandii*, which indicate that *NifEN* can complement when *vnfEN* are inactivated (Wolfinger & Bishop, 1991), a finding that is consistent with the fact that *vnfEN* in *A. vinelandii* is the result of a recent gene duplication of *nifEN* (Fig. 2).

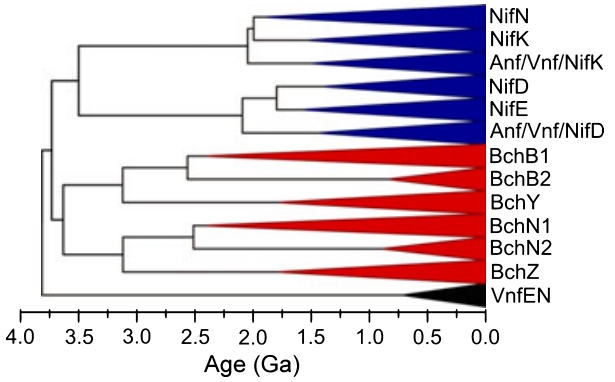

Fig. 4 Penalized-likelihood rate-smoothed chronogram of representative sequences for Vnf/Anf/NifDKEN and BchNBYZ rooted with representatives from the VnfEN lineage. Nitrogenase and Bch protein lineages are shaded in grey and black respectively. An uncollapsed version of the rate-smoothed chronogram is presented in Fig. S3.

Temporal relationship between the emergence of Nif and oxygenic photosynthesis

It has been previously suggested that Mo-nitrogenase proliferated in response to a rise in marine Mo concentrations that accompanied the GOE (Anbar & Knoll, 2002). However, it remains unclear if the Mo-nitrogenase evolved after the GOE and the increased availability of Mo, or if it emerged prior to the GOE when Mo is thought to have been in limited supply (Anbar *et al.*, 2007). To examine the timing of the emergence of Nif in the genomic record relative to the timing of the GOE and the rise of Mo availability in the geologic record, we investigated the evolutionary relationships between proteins involved in nitrogenase and bacteriochlorophyll (Bch) biosynthesis. If Mo-nitrogenase did in fact emerge as a result of widespread environmental oxygenation, then one would expect that the timing of the duplication event that yielded *nifEN* and the ability to incorporate Mo into the Nif active site cofactor would postdate the divergence of oxygenic and anoxygenic phototrophs in a phylogram containing Vnf/Anf/NifDKEN and BchNBYZ. Alternatively, if Mo-nitrogenase originated prior to the emergence of oxygenic photosynthesis and the potential for widespread environmental oxygenation, then the duplication event that yielded *nifEN* would predate the divergence of oxygenic and anoxygenic phototrophs in this phylogram.

To evaluate the relative plausibility of these two scenarios, we reconstructed the evolutionary history of representative Vnf/Anf/NifDKEN and BchNBYZ. The Vnf/Anf/NifDKEN and BchNBYZ phylogram reveals well-supported and distinct sublineages that, with the exception of 'VnfEN', cluster within either the nitrogenase or Bch stem lineages of the tree (Figs 4, S2 and S3). 'VnfEN' proteins branch near the root of the tree and between the nodes demarcating the nitrogenase and Bch lineage (Figs 4, S2 and S3), indicating that these proteins emerged very early with respect to proteins required for Mo-nitrogenase and photosynthesis. Two sublineages are present in both BchN and BchB clades (Figs 4, S2 and S3), all of which contain proteins derived from oxygenic and anoxygenic phototrophs (Fig. S3). In each of the four BchNB subclusters, proteins from anoxygenic phototrophs branch early with respect to oxygenic phototrophs, a finding that is consistent with other analyses of proteins involved in photosynthesis (Burke *et al.*, 1993; Xiong *et al.*, 2000).

To compare the ages of the four crown clades relevant to our question on the origin of Mo-nitrogenase (i.e. Vnf/Anf/NifDE, Vnf/Anf/NifKN, BchN and BchB), the phylogram was rate smoothed using three independent approaches. All three rate estimation approaches (PL, rate constancy and NP) yielded similar age estimates. Since derivations of age estimates from substitution rates centres on rate inconstancy, the fact that the three independent methods produce similar age estimates signals that the sequences are evolving closer to rate constancy. Using the PL algorithm with the root age set to 3.5 Ga and the age of the emergence of oxygenic photosynthesis set to 2.5 Ga, the estimated age of the Anf/Vnf/NifDE crown clade was 2.07 ± 0.10 Ga and the estimated age of the Vnf/Anf/NifKN crown clade was 1.84 ± 0.20 Ga (Fig. 4, Table 1). The estimated age of the NifDE crown clade (2.07 ± 0.10 Ga) and the NifKN crown clade (1.84 ± 0.20 Ga) was similar, providing further support for an in-tandem duplication of *nifDK* to *nifEN* at approximately 1.8–2.1 Ga using these specific time constraints. Importantly, additional computations using various other time constraints did not result in significant differences in the age estimates for nitrogenase crown clades (Table 1). For example, when the root age was set to 3.8 Ga and the age of the node demarcating the emergence of oxygenic photosynthesis was set to 2.8 Ga, the estimated age of the Anf/Vnf/NifDE crown clade increased slightly to 2.11 ± 0.07 Ga and the estimated age of the Vnf/Anf/NifKN crown clade increased slightly to 1.88 ± 0.20 Ga, when compared to ages

Table 1 Age estimates for nitrogenase crown clades as a function of varying root and oxygenic (O_2) photosynthesis fixed age constraints

Age constraints (Ga)		Nitrogenase crown clade ages (Ga)		
Root age	O_2 photosynthesis	Anf/Vnf/ NifDE	Anf/Vnf/ NifKN	Anf/Vnf/ NifDKEN
3.80	2.50	2.07 ± 0.07	1.84 ± 0.20	3.43 ± 0.10
3.80	2.80	2.11 ± 0.07	1.88 ± 0.20	3.46 ± 0.09
3.50	2.50	1.89 ± 0.06	1.68 ± 0.18	3.10 ± 0.08
3.50	2.80	1.93 ± 0.06	1.72 ± 0.18	3.12 ± 0.08

determined when the root age was set to 3.8 Ga and the age of the emergence of oxygenic photosynthesis was set to 2.5 Ga (Table 1). Similar observations were noted when iterations were run with other time constraints (Table 1). After incorporating the age estimates from all substitution rate estimate iterations using the various time constraints, the boundaries on the timing of the duplication of *nifDK* to *nifEN* can be estimated to be 1.5–2.2 Ga. When compared to the inferred timing of the node demarcating the duplication event that gave rise to the D and K lineages (crown clade age of Anf/Vnf/NifDKEN) at approximately 3.0–3.5 Ga (Table 1), this finding suggests that the duplication of *nifDK* to *nifEN* at approximately 1.5–2.2 Ga resulting in an enzyme complex (Mo-nitrogenase) that had superior properties to that of ancestral Vnf/Anf/NifDK enzymes synthesized in the absence of a scaffold complex (NifEN) as evinced by a lengthy gap in the extant evolutionary record of DK from approximately 3.0 to approximately 2.2 Ga (Fig. 4). Such an observation is consistent with biochemical and genetic studies which indicate that in the absence of the NifEN scaffold complex, the NifB-co precursor cannot be further matured to FeMo-co, resulting in a NifB-co/DK complex that could catalyze the reduction of a wide range of substrates but likely had poor reactivity towards N_2 (Curatti *et al.*, 2006; Rubio & Ludden, 2008; Soboh *et al.*, 2010).

The ages of NifDE and NifKN clades are each estimated at about 1.5–2.2 Ga, suggesting that the timing of the duplication of *nifDK* to *nifEN* and the ability to incorporate Mo into the nitrogenase co-factor just postdates the emergence of oxygenic photosynthesis at approximately 2.5–2.8 Ga (Canfield *et al.*, 2000; Holland, 2002; Anbar *et al.*, 2007). An origin for Mo-nitrogenase that follows the emergence of oxygenic photosynthesis is consistent with the distribution of *nifDKEN* in extant genomes and with phylogenetic reconstructions of NifDKEN, which collectively suggest that Nif evolved relatively recently and after the split between bacteria and archaea. Intriguingly, an emergence for Mo-nitrogenase at approximately 1.5–2.2 Ga is consistent with the hypothesized nitrogen crisis estimated at 2.2 Ga (Navarro-González *et al.*, 2001). Likewise, this estimated origin for Mo-nitrogenase at approximately 1.5–2.2 Ga is consistent with geochemical measurements which indicate a rise in Mo concentrations in the oceans just preceding this time approximately 2.5 Ga (Scott *et al.*, 2008), presumably due to increased weathering of continental sulphides in the presence of an increasingly oxidizing environment. This rise may have begun sporadically during the period 2.5–2.0 Ga, when Mo was clearly dissolved in seawater although at concentrations generally a hundred-fold or more lower than today (Anbar *et al.*, 2007). The increase in Mo availability over this time frame, in response to increasing oxygenation of the biosphere (Anbar *et al.*, 2007; Kaufman *et al.*, 2007; Scott *et al.*, 2008; Reinhard *et al.*, 2009; Kendall *et al.*, 2010) could have provided the evolutionary impetus to evolve the Mo-containing form of nitrogenase from ancestral enzymes that were unlikely to contain Mo at their active site and that were likely to exhibit poor reactivity towards N_2 (Curatti *et al.*, 2006; Rubio & Ludden, 2008; Soboh *et al.*, 2010).

However, while the phylogenetic data indicate that Mo-nitrogenase probably evolved after the rise of O_2 and the resulting increase in bioavailable Mo, the local microenvironment where Mo-nitrogenase evolved and proliferated was likely to be anoxic. This follows from the fact that the earliest branching NifDKEN proteins were from hydrogenotrophic methanogens which are obligate anaerobes (Boone *et al.*, 1993; Thauer *et al.*, 2008). But how could a nitrogenase with Mo as a constituent of its active site cluster evolve and proliferate in an anoxic environment, given the insoluble nature of Mo under reduced conditions which are presumed to characterize any environment co-inhabited by hydrogenotrophic methanogens and anaerobic *Firmicutes*? It is likely that this proliferation occurred during the earliest stages of environmental oxygenation, before the oceans became thoroughly oxygenated and where Mo was at least episodically available [e.g. anoxic, but non-sulphidic conditions (Helz *et al.*, 1996)]. A likely setting was a redox-stratified ocean basin, in which mildly oxygenated Mo-containing surface waters overlay deeper waters that were depleted in O_2 and H_2S, but which may have been rich in the reductant Fe^{2+}. The dissolved ferrous iron may have helped to maintain anoxic conditions at ocean depth by reacting with O_2 (Weber *et al.*, 2006; Slack *et al.*, 2007) that diffuses in from surface waters and or reacting with H_2S that may diffuse in from continental shelf environments that may otherwise precipitate Mo, rendering it unavailable to biology (Helz *et al.*, 1996). During the early Proterozoic (*c.* 1.5–2.2 Ga) when Mo-nitrogenase is thought to have originated, evidence indicates that deep ocean waters may have been depleted in SO_4^{2-} and H_2S, but were likely rich in Fe^{2+}, especially in water column environments extending >100 km beyond the continental shelf (Poulton *et al.*, 2010). In the absence of significant concentrations of SO_4^{2-} (<60 μM), hydrogenotrophic methanogens can outcompete sulphate-reducing bacteria (SRB) for the growth substrate H_2 (Winfrey & Zeikus, 1977; Lovley & Klug, 1983). Therefore, the effects of low concentrations of sulphate in a redox-stratified ocean basin may have promoted the emergence of Mo-nitrogenase within the hydrogenotrophic methanogens in

two ways. First, low concentrations of sulphate would favour the activity of hydrogenotrophic methanogens at the expense of SRB, due to competition for the growth substrate H_2 (Winfrey & Zeikus, 1977; Lovley & Klug, 1983). Secondly, the decreased SRB activity would preclude significant contributions of biologically sourced H_2S capable of precipitating episodic flux of Mo from overlaying waters (Helz *et al.*, 1996), which together would create a niche that would presumably have been prime for hydrogenotrophic methanogens to persist and evolve Mo-nitrogenase.

CONCLUSIONS

In the present study, we focused on key events (e.g. gene duplications, gene fusions) in the evolution of proteins that are required to biosynthesize the FeMo active site co-factor of Mo-nitrogenase to generate new insight into the evolution of biological nitrogen fixation and its temporal relationship with the evolution of oxygenic photosynthesis. Phylogenetic reconstruction of Anf/Vnf/NifDKEN sequences indicates that NifEN are nested among NifDK sequences, suggesting that *nifDK* duplicated and gave rise to *nifEN*. This observation, coupled with the finding that NifEN from archaea and bacteria are not reciprocally monophyletic, provide strong support for the hypothesis that Mo-nitrogenase was not present in LUCA and that it is a relatively recent evolutionary innovation. The phylogenetic branching order of NifDKEN protein sequences indicates that the duplication of *nifDK* to *nifEN*, and the ability to further mature NifB-co to FeMo-co, most likely occurred in an ancestor of the hydrogenotrophic methanogens, with acquisition in the bacterial domain via LGT between an ancestor of the methanogenic archaea and an ancestor of the *Firmicutes*. A comparison of substitution rates in Anf/Vnf/NifDKEN protein sequences and substitution rates of BchNBYZ protein sequences suggests that Nif emerged approximately 1.5–2.2 Ga, after the origin of oxygenic photosynthesis and the widespread oxygenation of the biosphere. Given the anaerobic physiologies of members of the methanogenic archaea and the *Firmicutes*, it is likely that the origin and early proliferation of Nif occurred in an anoxic environment where Mo is at least episodically available. A likely setting was a redox-stratified ocean basin, in which mildly oxygenated Mo-containing surface waters overlay waters that were depleted in O_2 and H_2S; conditions which are thought to have been common during the early Proterozoic in marine environments extending >100 km from the palaeoshoreline (Kendall *et al.*, 2010; Poulton *et al.*, 2010). This redox-stratified ocean basin would presumably facilitate episodic flux of Mo into deeper anoxic waters where hydrogenotrophic methanogens thrived at a time when abiotic sources of fixed N are thought to have become limiting (Navarro-González *et al.*, 2001), creating the complex set of conditions that may have precipitated the innovation of Mo-nitrogenase.

ACKNOWLEDGMENTS

We are grateful to Robert Blankenship, Donald Bryant and Luis Rubio for stimulating discussions and/or critical evaluations of the manuscript. We thank Aubrey Zerkle and two anonymous referees whose comments and suggestions greatly improved this manuscript. This work was supported by the NASA Astrobiology Institute (J.W.P and A.D.A.) and the National Science Foundation (S.R.M.). E.S.B. was supported by a NASA Astrobiology Institute postdoctoral fellowship and T.L.H. is supported by a National Science Foundation – Integrative Graduate Education and Research Training fellowship.

REFERENCES

Abascal F, Zardoya R, Posada D (2005) ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* **2**, 2104–2105.

Anbar AD (2008) Elements and evolution. *Science* **322**, 1481–1483.

Anbar AD, Knoll AH (2002) Proterozoic ocean chemistry and evolution: a bioinorganic bridge? *Science* **297**, 1137–1142.

Anbar AD, Duan Y, Lyons TW, Arnold GL, Kendall B, Creaser RA, Kaufman AJ, Gordon GW, Scott C, Garvin J, Buick R (2007) A whiff of oxygen before the great oxidation event? *Science* **317**, 1903–1906.

Baldauf SL, Palmer JD, Doolittle WF (1996) The root of the universal tree and the origin of eukaryotes based on elongation factor phylogeny. *Proceedings of the National Academy of Sciences of the USA* **93**, 7749–7754.

Beaumont V, Robert F (1999) Nitrogen isotope ratios of keragens in Precambrian cherts: a record of the evolution of atmospheric chemistry? *Precambrian Research* **96**, 63–82.

Beiko RG, Harlow TJ, Ragan MA (2005) Highways of gene sharing in prokaryotes. *Proceedings of the National Academy of Sciences of the USA* **102**, 14332–14337.

Boone DR, Whitman WB, Rouvière P (1993) Diversity and taxonomy of methanogens. In *Methanogenesis* (ed. Ferry JG). Chapman & Hall, New York/London, pp. 35–80.

Brigle KE, Weiss MC, Newton WE, Dean DR (1987) Products of the iron-molybdenum cofactor-specific biosynthetic genes, *nifE* and *nifN*, are structurally homologous to the products of the nitrogenase molybdenum-iron protein genes, *nifD* and *nifK*. *Journal of Bacteriology* **169**, 1547–1553.

Brocks JJ, Buick R, Logan GA, Summons RE (2003) Composition and syngeneity of molecular fossils from the 2.78 to 2.45 billion-year-old Mount Bruce Supergroup, Pilbara Craton, Western Australia. *Geochimica et Cosmochimica Acta* **67**, 4289–4319.

Bulen WA, Lecomte JR (1966) The nitrogenase system from *Azotobacter*: two enzyme requirements for N_2 reduction, ATP dependent H_2 evolution and ATP hydrolysis. *Proceedings of the National Academy of Sciences of the USA* **56**, 979–986.

Burke DH, Hearst JE, Sidow A (1993) Early evolution of photosynthesis: clues from nitrogenase and chlorophyll iron proteins. *Proceedings of the National Academy of Sciences of the USA* **90**, 7134–7138.

Canfield DE, Habicht KS, Thamdrup B (2000) The archean sulfur cycle and the early history of atmospheric oxygen. *Science* **288**, 658–661.

Chew AGM, Bryant DA (2007) Chlorophyll biosynthesis in bacteria: the origins of structural and functional diversity. *Annual Review of Microbiology* **61**, 113–129.

Christiansen J, Goodwin PJ, Lanzilotta WN, Seefeldt LC, Dean DR (1998) Catalytic and biophysical properties of a nitrogenase Apo-MoFe protein produced by a *nifB*-deletion mutant of *Azotobacter vinelandii*. *Biochemistry* **37**, 12611–12623.

Curatti L, Ludden PW, Rubio LM (2006) NifB-dependent *in vitro* synthesis of the iron-molybdenum cofactor of nitrogenase. *Proceedings of the National Academy of Sciences of the USA* **103**, 5297–5301.

Eady RR (1996) Structure-function relationships of alternative nitrogenases. *Chemical Reviews* **96**, 3013–3030.

Eigenbrode JL, Freeman KH, Summons RE (2008) Methylhopane biomarker hydrocarbons in Hamersley Province sediments provide evidence for Neoarchean aerobiosis. *Earth and Planetary Science Letters* **273**, 323–331.

Erixon P, Svennblad B, Britton T, Oxelman B (2003) Reliability of Bayesian posterior probabilities and bootstrap frequencies in phylogenetics. *Systematic Biology* **52**, 665–673.

Fani R, Gallo R, Liò P (2000) Molecular evolution of nitrogen fixation: the evolutionary history of the *nifD*, *nifK*, *nifE*, and *nifN* genes. *Journal of Molecular Evolution* **51**, 1–11.

Fournier GP, Gogarten JP (2008) Evolution of acetoclastic methanogenesis in *Methanosarcina* via horizontal gene transfer from cellulolytic *Clostridia*. *Journal of Bacteriology* **190**, 1124–1127.

Fujita Y, Matsumoto H, Takahashi Y, Matsubara H (1993) Identification of a *nifDK*-like gene (ORF467) involved in the biosynthesis of chlorophyll in the cyanobacterium *Plectonema boryanum*. *Plant and Cell Physiology* **34**, 305–314.

Garvin J, Buick R, Anbar AD, Arnold GL, Kaufman AJ (2009) Isotopic evidence for an aerobic nitrogen cycle in the latest Archean. *Science* **323**, 1045–1048.

Godfrey LV, Falkowski PG (2009) The cycling and redox state of nitrogen in the Archaean ocean. *Nature Geoscience* **2**, 725–729.

Hartmann LS, Barnum SR (2010) Inferring the evolutionary history of Mo-dependent nitrogen fixation from phylogenetic studies of *nifK* and *nifDK*. *Journal of Molecular Evolution* **71**, 70–85.

Hearst JE, Alberti M, Doolittle RF (1985) A putative nitrogenase reductase gene found in the nucleotide sequences from the photosynthetic gene cluster of *R. capsulata*. *Cell* **40**, 219–220.

Helz GR, Miller CV, Charnock JM, Mosselmans JFW, Pattrick RAD, Garner CD, Vaughan DJ (1996) Mechanism of molybdenum removal from the sea and its concentration in black shales: EXAFS evidence. *Geochimica et Cosmochimica Acta* **60**, 3631–3642.

Holland HD (2002) Volcanic gases, black smokers, and the Great Oxidation Event. *Geochimica et Cosmochimica Acta* **66**, 3811–3826.

Hu Y, Fay AW, Ribbe MW (2005) Identification of a nitrogenase FeMo cofactor precursor on NifEN complex. *Proceedings of the National Academy of Sciences of the USA* **102**, 3236–3241.

Hu Y, Yoshizawa JM, Fay AW, Lee CC, Wiig JA, Ribbe MW (2009) Catalytic activities of NifEN: implications for nitrogenase evolution and mechanism. *Proceedings of the National Academy of Sciences of the USA* **106**, 16962–16966.

Hu Y, Fay AW, Lee CC, Wiig JA, Ribbe MW (2010) Dual functions of NifEN: insights into the evolution and mechanism of nitrogenase. *Dalton Transactions* **39**, 2964–2971.

Huelsenbeck JP, Ronquist R (2001) MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* **17**, 754–755.

Jacobson MR, Brigle KE, Bennett LT, Setterquist RA, Wilson MS, Cash VL, Beynon J, Newton WE, Dean DR (1989) Physical and genetic map of the major nif gene cluster from *Azotobacter vinelandii*. *Journal of Bacteriology* **171**, 1017–1027.

Joerger RD, Bishop PE (1988) Bacterial alternative nitrogen fixation systems. *Critical Reviews in Microbiology* **16**, 1–14.

Kaiser JT, Hu Y, Wiig JA, Rees DC, Ribbe MW (2011) Structure of precursor-bound NifEN: a nitrogenase FeMo cofactor maturase/insertase. *Science* **331**, 91–94.

Kasting JF, Siefert JL (2001) Biogeochemistry: the nitrogen fix. *Nature* **412**, 26–27.

Kasting JF, Walker JCG (1981) Limits on oxygen concentrations in the prebiological atmosphere and the rate of abiotic nitrogen fixation. *Journal of Geophysical Research* **86**, 1147–1158.

Kaufman AJ, Johnston DT, Farquhar J, Masterson AL, Lyons TW, Bates S, Anbar AD, Arnold GL, Garvin J, Buick R (2007) Late Archean biospheric oxygenation and atmospheric evolution. *Science* **317**, 1900–1903.

Kendall B, Reinhard CT, Lyons TW, Kaufman AJ, Poulton SW, Anbar AD (2010) Pervasive oxygenation along late Archaean ocean margins. *Nature Geoscience* **3**, 647–652.

Kessler PS, Mclarnan J, Leigh JA (1997) Nitrogenase phylogeny and the molybdenum dependence of nitrogen fixation in *Methanococcus maripaludis*. *Journal of Bacteriology* **179**, 541–543.

Kim J, Rees DC (1992) Crystallographic structure and functional implications of the nitrogenase molybdenum-iron protein from *Azotobacter vinelandii*. *Nature* **360**, 553–560.

Larkin MA, Blackshields G, Brown NP, Chenna R, Mcgettigan PA, Mcwilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948.

Leigh JA (2000) Nitrogen fixation in methanogens: the archaeal perspective. *Current Issues in Molecular Biology* **2**, 125–131.

Lovley DR, Klug MJ (1983) Sulfate reducers can outcompete methanogens at freshwater sulfate concentrations. *Applied and Environmental Microbiology* **45**, 187–192.

Mojzsis SJ, Arrhenius G, Mckeegan KD, Harrison TM, Nutman AP, Friend CRL (1996) Evidence for life on Earth before 3,800 million years ago. *Nature* **384**, 55–59.

Navarro-González R, Mckay CP, Mvondo DN (2001) A possible nitrogen crisis for Archaean life due to reduced nitrogen fixation by lightning. *Nature* **412**, 61–64.

Page RDM (1996) TREEVIEW: an application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* **12**, 357–358.

Pagel M (1999) Inferring the historical patterns of biological evolution. *Nature* **401**, 877–884.

Parsons I, Lee MR, Smith JV (1998) Biochemical evolution II: origin of life in tubular microstructures on weathered feldspar surfaces. *Proceedings of the National Academy of Sciences of the USA* **95**, 15173–15176.

Penny D, Poole A (1999) The nature of the last universal common ancestor. *Current Opinion in Genetics & Development* **9**, 672–677.

Peters JW, Fisher K, Dean DR (1995) Nitrogenase structure and function: a biochemical-genetic perspective. *Annual Review of Microbiology* **49**, 335–366.

Poulton SW, Fralick P, Canfield DE (2010) Spatial variability in oceanic redox structure 1.8 billion years ago. *Nature Geoscience* **3**, 486–490.

Raymond J, Siefert JL, Staples CR, Blankenship RE (2004) The natural history of nitrogen fixation. *Molecular Biology and Evolution* **21**, 541–554.

Reinhard CT, Raiswell R, Scott C, Anbar AD, Lyons TW (2009) A late Archean sulfidic sea stimulated by early oxidative weathering of the continents. *Science* **326**, 713–716.

Rodriguez-Trelles F, Tarrio R, Ayala FJ (2002) A methodological bias toward overestimation of molecular evolutionary time scales. *Proceedings of the National Academy of Sciences of the USA* **12**, 8112–8115.

Roll JT, Shah VK, Dean DR, Roberts GP (1995) Characteristics of NIFNE in *Azotobacter vinelandii* strains. *Journal of Biological Chemistry* **270**, 4432–4437.
Ronquist F, Huelsenbeck JP (2003) MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574.
Rubio LM, Ludden PW (2005) Maturation of nitrogenase: a biochemical puzzle. *Journal of Bacteriology* **187**, 405–414.
Rubio LM, Ludden PW (2008) Biosynthesis of the iron-molybdenum cofactor of nitrogenase. *Annual Review of Microbiology* **62**, 93–111.
Sanderson MJ (1997) A nonparametric approach to estimating divergence times in the absence of rate constancy. *Molecular Biology and Evolution* **14**, 1218–1231.
Sanderson MJ (2002) Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Molecular Biology and Evolution* **19**, 101–109.
Sanderson MJ (2004) r8s, version 1.70. User's manual.
Schopf JW (1992) Microfossils of the early Archean Apex Chert: new evidence of the antiquity of life. *Science* **260**, 640–646.
Scott C, Lyons TW, Bekker A, Shen Y, Poulton SW, Chu X, Anbar AD (2008) Tracing stepwise oxygenation of the Proterozoic ocean. *Nature* **452**, 456–459.
Shah VK, Allen JR, Spangler NJ, Ludden PW (1994) *In vitro* synthesis of the iron-molybdenum cofactor of nitrogenase. *Journal of Biological Chemistry* **269**, 1154–1158.
Slack JF, Grenne T, Bekker A, Rouxel OJ, Lindberg PA (2007) Suboxic deep seawater in the late Paleoproterozoic: evidence from hematitic chert and iron formation related to seafloor-hydrothermal sulfide deposits, central Arizona, USA. *Earth and Planetary Science Letters* **255**, 243–256.
Soboh B, Boyd ES, Zhao D, Peters JW, Rubio LM (2010) Substrate specificity and evolutionary implications of a NifDK enzyme carrying NifB-co at its active site. *FEBS Letters* **584**, 1487–1492.
Stams A (1994) Metabolic interactions between anaerobic bacteria in methanogenic environments. *Antonie van Leeuwenhoek* **66**, 271–294.
Summons RE, Jahnke LL, Hope JM, Logan GA (1999) 2-Methylhopanoids as biomarkers for cyanobacterial oxygenic photosynthesis. *Nature* **400**, 554–557.
Thauer RK, Kaster A-K, Seedorf H, Buckel W, Hedderich R (2008) Methanogenic archaea: ecologically relevant differences in energy conservation. *Nature Reviews. Microbiology* **6**, 579–591.
Thiel T (1996) Isolation and characterization of the VnfEN genes of the cyanobacterium *Anabaena variabilis*. *Journal of Bacteriology* **178**, 4493–4499.
Ugalde RA, Imperial J, Shah VK, Brill WJ (1984) Biosynthesis of iron-molybdenum cofactor in the absence of nitrogenase. *Journal of Bacteriology* **159**, 888–893.
Weber KA, Achenbach LA, Coates JD (2006) Microorganisms pumping iron: anaerobic microbial iron oxidation and reduction. *Nature Reviews. Microbiology* **4**, 752–764.
Winfrey MR, Zeikus JG (1977) Effect of sulfate on carbon and electron flow during microbial methanogenesis in freshwater sediments. *Applied and Environmental Microbiology* **33**, 275–281.
Wolfinger ED, Bishop PE (1991) Nucleotide sequence and mutational analysis of the vnfENX region of Azotobacter vinelandii. *Journal of Bacteriology* **173**, 7565–7572.
Xiong J, Fischer WM, Inoue K, Nakahara M, Bauer CE (2000) Molecular evidence for the early evolution of photosynthesis. *Science* **289**, 1724–1730.
Yung YL, Mcelroy MB (1979) Fixation of nitrogen in the prebiotic atmosphere. *Science* **203**, 1002–1004.
Zehr JP, Jenkins BD, Short SM, Steward GF (2003) Nitrogenase gene diversity and microbial community structure: a cross-system comparison. *Environmental Microbiology* **5**, 539–554.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Fig. S1. Bayesian consensus phylogenetic tree of Nif/Vnf/AnfDK and BchNBYZ proteins from taxa representing the primary lineages for each locus.

Fig. S2. Bayesian consensus phylogenetic tree of Vnf/Anf/NifDKEN and BchNBYZ representative loci.

Fig. S3. Rate-smoothed radial chronogram of representative Vnf/Anf/NifDKEN and BchNBYZ loci as estimated using the penalized likelihood method.

Table S1. Accession numbers for nitrogenase protein sequences used in this study.

Table S2. Accession numbers for BchNBYZ loci used in this study.

Please note: Wiley–Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.