The Natural History of Nitrogen Fixation

Jason Raymond,* Janet L. Siefert,† Christopher R. Staples,* and Robert E. Blankenship*

*Department of Chemistry and Biochemistry, Arizona State University, Tempe; and †Department of Statistics, Rice University, Houston

In recent years, our understanding of biological nitrogen fixation has been bolstered by a diverse array of scientific techniques. Still, the origin and extant distribution of nitrogen fixation has been perplexing from a phylogenetic perspective, largely because of factors that confound molecular phylogeny such as sequence divergence, paralogy, and horizontal gene transfer. Here, we make use of 110 publicly available complete genome sequences to understand how the core components of nitrogenase, including NifH, NifD, NifK, NifE, and NifN proteins, have evolved. These genes are universal in nitrogen fixing organisms—typically found within highly conserved operons—and, overall, have remarkably congruent phylogenetic histories. Additional clues to the early origins of this system are available from two distinct clades of nitrogenase paralogs: a group composed of genes essential to photosynthetic pigment biosynthesis and a group of uncharacterized genes present in methanogens and in some photosynthetic bacteria. We explore the complex genetic history of the nitrogenase family, which is replete with gene duplication, recruitment, fusion, and horizontal gene transfer and discuss these events in light of the hypothesized presence of nitrogenase in the last common ancestor of modern organisms, as well as the additional possibility that nitrogen fixation might have evolved later, perhaps in methanogenic archaea, and was subsequently transferred into the bacterial domain.

Introduction

Biologically available nitrogen, also called fixed nitrogen, is essential for life. All known nitrogen-fixing organisms (diazatrophs) are prokaryotes, and the ability to fix nitrogen is widely, though paraphyletically, distributed across both the bacterial and archaeal domains (fig. 1). The capacity for nitrogen fixation in these organisms relies solely upon the nitrogenase enzyme system, which, at 16 ATPs hydrolyzed per N_2 fixed, carries out one of the most metabolically expensive processes in biology (Simpson and Burris 1984). The amount of biologically fixed nitrogen produced today is in excess of 2×10^{13} g/year (Falkowski 1997). In contrast, lightning discharge—the primary abiotic source of fixed nitrogen—accounts for 10^{12} to 10^{13} g/year. It has been suggested that abiotic sources of fixed nitrogen on the early Earth, supplied through endogenous/abiotic synthesis and exogenous delivery, were most likely limiting (Raven and Yin 1998; Kasting and Siefert 2001; Navarro-Gonzalez, McKay, and Mvondo 2001). At some point, dwindling concentrations of reduced nitrogen would have become insufficient for an expanding microbial biomass, precipitating the evolution of biological nitrogen fixation (Towe 2002). These prevailing conditions have been used to argue that the innovation of biological nitrogen fixation occurred early in prokaryotic evolution, and indeed the ability to fix nitrogen is found exclusively among members of the bacteria and archaea.

Our understanding of the contemporary nitrogenase system as it has evolved in various lineages is contingent upon understanding the genomic events that produced and stabilized the genetic system and operon(s) structure, either through vertical or horizontal transmission. This enzyme system has been clearly mobile through nonvertical transmission. There are numerous instances of nitrogenase genes and operons being selectively lost, duplicated, horizontally transferred, and, in at least one significant case, recruited into the photosynthetic apparatus pathway (Xiong et al. 2000). Occurrences of nitrogen fixation (*nif*) gene fusions have been reported in several organisms, as has the presence of *nif* genes on plasmids (Prakash, Schilperoort, and Nuti 1981; Thiel 1993; Goodman and Weisz 2002). In almost all cases, *nif* genes are found within one or several extensive, cotranscribed operons or regulons that not only encode the subunits of the functional nitrogenase protein but also code for an expansive suite of proteins involved with regulation, activation, metal transport, and cluster biosynthesis (Kessler, Blank, and Leigh 1998; Kessler and Leigh 1999; Halbleib and Ludden 2000).

To understand the evolutionary history of nitrogenase, we must also consider any additional selective pressures that affect the stability and distribution of diazatrophy. Nitrogen fixation occurs in a varied metabolic context in both aerobic and anaerobic environments; yet nitrogenase is irreversibly inactivated by oxygen in vitro (although reversible oxidation has been observed in vivo [Zehr et al. 1993]). Maintaining the ability to fix nitrogen in the presence of exogenous or endogenous sources of O_2 has necessitated innovative biochemical and physiological mechanisms for segregation. This is particularly important in oxygen-evolving cyanobacteria and organisms carrying out oxidative phosphorylation. Certain cyanobacteria contribute substantial amounts of fixed nitrogen in marine environments and do so because of exquisite controls on temporal and spatial separation of the two processes (Berman-Frank et al. 2001).

Nitrogenase itself is an ATP-hydrolyzing, redox-active complex of two component proteins, the dinitrogenase $\alpha_2\beta_2$ heterotetramer (where $\alpha = $ NifD and $\beta = $ NifK proteins) and the dinitrogenase reductase γ_2 homodimer (NifH protein). The α subunit contains the active site for dinitrogen reduction, typically a $MoFe_7S_9$ metal cluster (termed FeMo-cofactor), although some organisms contain nitrogenases wherein Mo is replaced by either Fe or V (in which case the nomenclature Anf or Vnf, respectively, is used instead of Nif). These so-called alternative
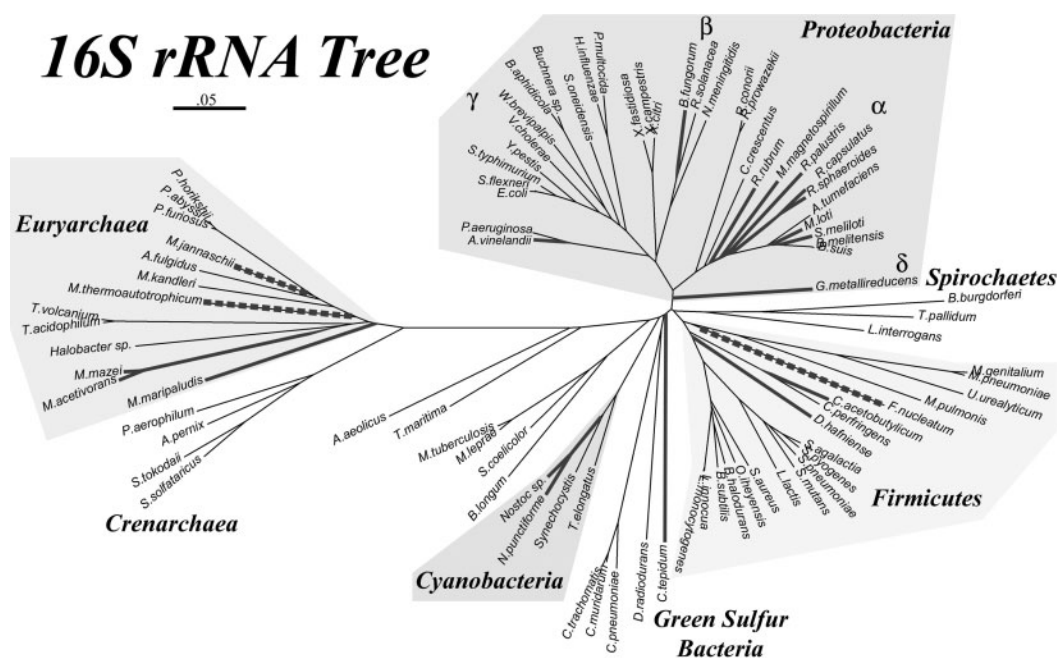
Fig. 1.—Neighbor-joining phylogenetic tree constructed from 16S rDNA sequences for the microbial genomes used in this analysis. Diazotrophic genomes, as determined by the presence of NifHDKEN operons, are indicated with bold lines. Lineages outlined by dashed bold lines have homologs to NifH and NifD but are not known to fix nitrogen. Also shown are the major bacterial and archaeal phyla, highlighted if nitrogen-fixing lineages are found among them.

nitrogenases are found only in a limited subset of diazotrophs and, in all cases studied so far, are present secondarily to the FeMo nitrogenase. Alternative nitrogenases are expressed only when Mo concentrations are limiting, with the Vnf nitrogenase being expressed preferentially to the Anf nitrogenase if V and Fe are both present and the organism possesses all three types of nitrogenases. The FeMo nitrogenase has been found to be more specific for and more efficient in binding dinitrogen and reducing it to ammonia than either of the alternative nitrogenases (in the order Nif > Vnf > Anf) (Joerger and Bishop 1988; Miller and Eady 1988). Although these alternative nitrogenases could be derived paralogs of the Mo-requiring enzyme, they may instead represent primitive nitrogenases that have been maintained in several diverse lineages of prokaryotes (Anbar and Knoll 2002).

Adjacent to the *nifH*, *nifD*, and *nifK* genes in most *nif* operons are the *nifE* and *nifN* genes, typically, although variably, in the order *nifHDKEN*. The NifE and NifN proteins have significant similarity to NifD and NifK, respectively, and are believed to have originated from an ancient duplication of a NifDK operon (Fani, Gallo, and Lio 2000). NifE and NifN are thought to function as scaffolds for FeMo-cofactor or FeV-cofactor assembly and are typically if not always associated with Mo-dependent nitrogenase operons (Roll et al. 1995). Alternative nitrogenase genes, particulary the Fe-only (*anf*) nitrogenases, are typically found without contiguous E or N homologs, suggesting differences in protein requirements for cofactor construction or the necessity for an assembly scaffold. All four of these proteins are distantly similar based on sequence alignment, suggesting a single ancestral nitrogenase homolog—perhaps originally functional as

a homodimer (Fani, Gallo, and Lio 2000)—that has given rise to the modern family of NifD homologs.

Our understanding of the phylogenetic history of *nif* genes has been advanced significantly by extensive efforts in sequencing nitrogenase genes, primarily the highly conserved *nifH* gene but also the larger but less conserved *nifD*, *nifK*, *nifE*, and *nifN* genes (Normand and Bousquet 1989; Normand et al. 1992; Hirsch et al. 1995; Zehr, Mellon, and Hiorns 1997; Fani, Gallo, and Lio 2000). Because of the inherent difficulties of obtaining extended fragments of DNA, the evolution of the nitrogenase operon has been necessarily considered piecemeal, typically through analysis of NifH protein sequences. The presence of multiple paralogous nitrogenase gene copies within single genomes has the potential to confound single gene analyses. In an attempt to circumvent this limitation, we have accumulated sequence data as well as information on operon structure and gene synteny for *nifHDKEN* homologs in all of the currently available genomes from the NCBI sequence repository. The current database of prokaryote genomes includes representatives that span the known phylogenetic *nif* gene groups, and additional sequences from NCBI were included in the analysis to increase taxonomic breadth. Using these data, we have been able to infer phylogenies for the major genes associated with the *nif* operon in attempt to understand the timing and complex genetic events that have marked the history of nitrogen fixation.

Materials and Methods

Annotated protein sequences for all completed and publicly available genomes were downloaded from NCBI (http://www.ncbi.nlm.nih.gov; 101 genomes total when the

analysis was undertaken). Additional nearly complete genomes were downloaded from the Joint Genome Institute Web site (http://www.jgi.doe.gov/), and the *Rhodobacter capsulatus* genome was downloaded from the Integrated Genomics Web site (see Supplementary Material online at www.mbe.oupjournals.org for specific genomes and database). All genomes were screened for NifHDKEN homologs and selected for further analysis if any BlastP alignment (Altschul et al. 1990) against a local database of Nif-like proteins (NifHDKEN proteins from *Azotobacter vinelandii* and *Clostridium acetobutylicum* and homologous pigment biosynthesis genes from *Rhodobacter sphaeroides*) yielded an expectation value (e value) of 1×10^{-10} or less. Using BlastP, this subset was then screened for presence of *nif* operons by searching for contiguous regions of the genome with significant similarity (Blast e value of 1×10^{-10} or lower) to a large set of nitrogen fixation protein sequences (including NifH, NifD, NifK, NifE, NifN, NifX, NifU, NifS, NifT, NifZ, NifW, NifV, and NifB; AnfH, AnfD, AnfK, AnfA, and AnfG; and VnfH, VnfD, VnfK, VnfA, and VnfG), retaining information on homolog length, position within the genome, most similar Nif protein and the strength of that alignment.

Individual sequences found to be homologous to known Nif proteins were then segregated into categories based on their closest Nif protein homolog, five of which (NifH, NifD, NifK, NifE, and NifN) are detailed herein. Fused nitrogenase (or Nif-homolog) proteins were found in the genomes of *C. acetobutylicum* (NifKB) and *M. acetivorans* (NifHD) and were retained in their respective analyses by dividing them into their most likely component proteins based on alignment (e.g., NifKB fusion split into NifK and NifB, for which only NifK is presented herein). Each of these categories was aligned using ClustalW version 1.81 (Thompson, Higgins, and Gibson 1994) with default parameters. Because of the high sequence divergence observed between nitrogenase and its distantly related homologs (less than 30% amino acid sequence identity for many pairwise comparisons; for example, among the large α and β subunits of the Mo-dependent versus alternative nitrogenases, which, despite their high divergence still perform analogous roles in fixing nitrogen [see the discussion of group III below]), phylogenies presented herein were reconstructed from amino acid sequences. Amino acid sequence alignments were constructed and phylogenies inferred using all available positions of characterized nitrogenases and protein sequences most likely to be functional nitrogenase subunits (i.e., groups I to III, as outlined below), as determined based on alignments, operon structure, conserved ligand-binding motifs, and their local phylogenetic group. This was done to avoid the pitfalls, such as long-branch attraction, associated with alignment and phylogenetic reconstruction involving highly diverged sequences. MEGA2 (Kumar, Tamura, and Nei 1994) was used to construct minimal-evolution (distance) trees (CNI search level of 2; gapped sites were pairwise deleted) using Poisson corrected distances and incorporating the alpha parameter for the eight-category gamma distribution as calculated with Tree-Puzzle version 5.01 (Schmidt et al. 2002) and bootstrapped 500 times. PHYLIP's ProML,

SeqBoot, and Consense (Felsenstein 1989) were used to infer and assemble bootstrapped maximum-likelihood trees (100 replicates), using the JTT model of amino acid substitution and also incorporating the Tree-Puzzle–estimated gamma distribution shape parameter to account for rate variation among sites. Phylogenies are presented below for NifH, NifD, and NifE; NifK and NifN are largely congruent to the latter two, respectively, and therefore are presented as supplemental figures. 16S rRNA–encoding DNA sequences were obtained from the whole genome sequences compared herein and aligned using the online tools and alignment masks available via the Ribosomal Database Project II (Cole et al. 2003).

Results and Discussion
Distinct Phylogenetic Clades

Early analyses of NifH sequences revealed distinct clusters of nitrogenase homologs, believed to be the modern paralogs resulting from multiple gene duplication events (Wang, Chen, and Johnson 1988; Normand and Bousquet 1989). The phylogenetic distribution of diazotrophs, according to a 16S rRNA–based species tree for completed microbial genomes, is shown in figure 1, illustrating the typically sporadic occurrence of nitrogen fixation among prokaryote families. Intriguingly, further phylogenetic analysis of each of the component nitrogenase proteins and their known homologs indicates that they segregate into distinct, topologically consistent clades as summarized in figure 2 (and shown in detail in figs 3, 4, and 5 for the NifH, D, and E proteins, respectively). These five groups are designated herein as (1) typical Mo-Fe nitrogenases, predominantly composed of members of the proteobacterial and cyanobacterial phyla; (2) anaerobic Mo-Fe nitrogenases from a wide range of predominantly anaerobic organisms, including clostridia, acetogenic bacteria, and several methanogens; (3) alternative nitrogenases, including the Mo-independent Anf and Vnf genes (except VnfH, which in all cases is more similar to NifH rather than AnfH); (4) uncharacterized *nif* homologs detected only in methanogens and some anoxygenic photosynthetic bacteria; and (5) bacteriochlorophyll and chlorophyll biosynthesis genes common to all phototrophs. The phylogenetic tree shown in figure 2, constructed from concatenated homologs of the NifH and NifD proteins, also supports the notion that the alternative nitrogenases are the earliest diverging of the three groups of true nitrogenases. The individual genes composing the core HDKEN operon have remarkably similar phylogenetic histories despite instances of gene duplication, rearrangement, and loss apparent in the records of multiple genomes. This suggests that nitrogenase gene family has evolved largely as a concerted unit, whose interaction and stoichiometry has been under purifying selection, thereby maintaining the highly conserved *nif* operon structure observed across the entire taxonomic spectrum of diazotrophs.

Group I

This clade consists primarily of Nif sequences from cyanobacteria and proteobacteria, which collectively
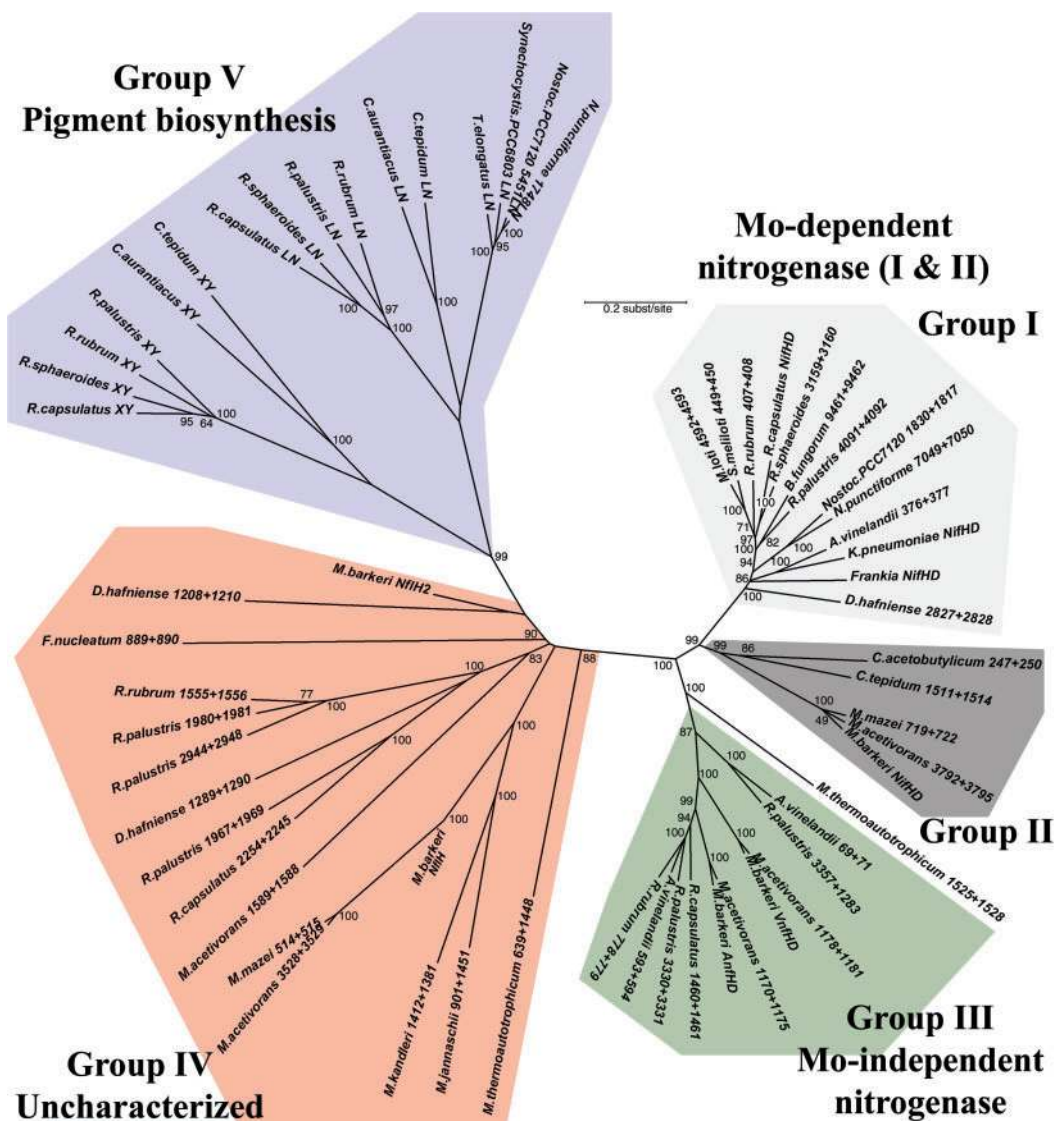
FIG. 2.—Overview of five phylogenetic groups elucidated in the text, shown on a concatenated phylogenetic tree composed of NifH and NifD homologs found in complete genomes. Groups I to III are all functional nitrogenases, including Mo-dependent group I and II nitrogenases as well as vanadium-dependent and iron-dependent group III homologs. Group V includes the subunits of protochlorophyllide reductase (here, BchX and Y concatenated) and chlorophyllide reductase (BchL and N concatenated) involved in the crucial late steps in photosynthetic pigment biosynthesis, which segregate into two distinct clades. Group IV consists of yet to be characterized nitrogenase paralogs of the NifH and NifD proteins, many of which are present in organisms not known to fix nitrogen. Because of significant sequence divergence, group IV homologs do not resolve to a single origin and so the branching within this group should be interpreted cautiously. The tree was constructed using the Neighbor-Joining method, as described in the text, with 500 bootstrap replicates (>60% bootstrap support percentages are shown at each node). Numbers after the organism names indicate the relative position of the relevant NifH and NifD homologs (first and second numbers, respectively) within that particular genome, and many organisms in fact have multiple sets of nitrogenase homologs and are thus found multiple times on the tree. Corresponding accession numbers are given in the Supplementary Material online, and all alignments are available via correspondence.

represent the best-studied nitrogenases and are among the largest *nif* gene operons. The genes comprising these extensive operons are involved mainly in nitrogenase regulation and assembly, not surprising given the physiological complexity characteristic of organisms from these phyla. Additionally, these two diverse bacterial groups are intimately associated with O_2 by way of aerobic respiration and oxygenic photosynthesis. Both phyla have intricate but well described spatial and/or temporal mechanisms for keeping nitrogenase and molecular oxygen separate, with

responsible genes often encoded within the *nif* operon (Adams 2000; Berman-Frank et al. 2001).

Although the group I topology is loosely consistent with 16S-based classification, several notable deviations exist. The α-proteobacteria form a nearly monophyletic clade within group I for each protein tree, with the recurring exception of *Rhodopseudomonas palustris* clustering alternatively with β-proteobacteria and γ-proteobacteria in NifDK and NifEN trees (figs. 4 and 5 and figs. S1 and S2 in Supplementary Material online at
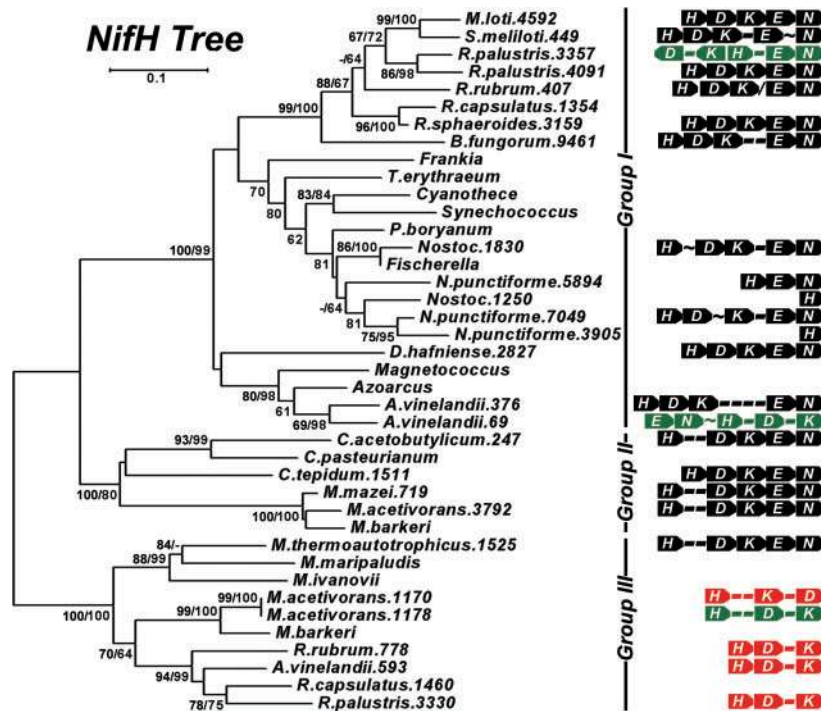
FIG. 3.—NifH phylogeny. Shown at the right of each figure is the position of the nitrogenase groups outlined in the text, as well as the local nitrogenase operon structure showing the relative positions of the HDKEN proteins (operons are color coded in black, green, or red for Mo-dependence, V-dependence, or Fe-dependence, respectively). For organisms with whole genome sequences available, the number after the organism name indicates the relative position of the protein's open reading frame within that genome (except in the cases of *R. capsulatus* and *R. palustris*). Within the operon structures, each dash (-) indicates a single open reading frame, the approximation symbol (~) denotes five to 10 open reading frames, and the slash (/) indicates that there is no clear local/operonal association between genes. Numbers at tree nodes indicate bootstrap support based on 100 maximum-likelihood replicates (first number) or 500 distance replicates (second number). The background tree is the distance tree; a short dash (-) at a node indicates low bootstrap support (<60%) for that branching pattern for either the ML tree or distance tree (although the same topology was observed). A single number or no number indicates a disagreement in the topology at that node between the ML and distance trees, with a single number indicating bootstrap support from the distance tree if it is 60% or greater. The scale bar gives the number of substitutions per site for branches on the distance tree.

www.mbe.oupjournals.org), respectively. The sporadic but typically well-supported branching of *R. palustris* proteins reflects several paralogous and perhaps horizontally transferred *nif* genes and operons present within its genome, possibly selected for by the diverse metabolic phenotypes this organism possesses (Imhoff 1995). A distinct separation between other proteobacterial classes is not obvious; for example, *Burkholderia fungorum* (β) and *R. palustris* (α) NifD, NifK, NifE, and NifN proteins cluster together (figures 4 and 5 and figures S1 and S2 in Supplementary Material online), and *Azoarcus* sp. (β), *A. vinelandii* (γ), and *Klebsiella pneumoniae* (γ) form a coherent branch where sequence data were available (figs. 3 and 4 and figs. S1 and S2 in Supplementary Material online; although notably in figs. 5 and S2, an additional pair of EN paralogs from a Vnf-associated cluster in *R. palustris* group with corresponding genes from *A. vinelandii*). Interestingly, the recurring *Azoarcus*, *A. vinelandii*, and *K. pneumoniae* group is supported by their shared unique two-component NifL/NifA regulatory system (Egener et al. 2002), corroborating the idea that gene sharing, perhaps involving entire *nif* regulons, has been promiscuous between these classes of organisms.

In the NifD and NifK trees (fig. 4 and fig. S1 in Supplementary Material online at www.mbe.oupjournals. org), the gram-positive bacteria *Frankia* sp. and *Desulfito-* *bacterium hafniense* both bridge the gap between group I and group II clades (the same position is observed for *D. hafniense* in NifE and NifN trees), in better agreement with their position on 16S trees. However, the *nifH* genes from both organisms are found within the group I clade, although with poor bootstrap support in the case of *D. hafniense*. The clustering of *Frankia*, an actinorhizal symbiont, with cyanobacterial phyla in the NifH phylogeny (fig. 3) is conspicuous and has been discussed previously (Normand et al. 1992; Hirsch et al. 1995). In several trees, the cyanobacteria cluster within proteobacterial classes, although the bootstrap support and consistency of this bifurcation is not robust. Furthermore, the presence of duplicate copies of NifH genes in the two *Nostoc* genomes compared here raises an important caveat in that comparisons relying on unsequenced genomes may be incomplete.

The mottled distribution of nitrogen fixation among cyanobacteria (*nif* genes have been found only in three of the six classified orders of cyanobacteria) may reflect the dilemma of carrying out both nitrogen fixation and oxygenic photosynthesis, because of the irreversible inactivation of nitrogenase by O_2. Diazotrophic cyanobacteria have evolved several mechanisms for segregating nitrogen fixation from O_2 (Adams 2000; Berman-Frank et al. 2001). Whether nitrogenase was present in the cyanobacterial ancestor or was horizontally transferred
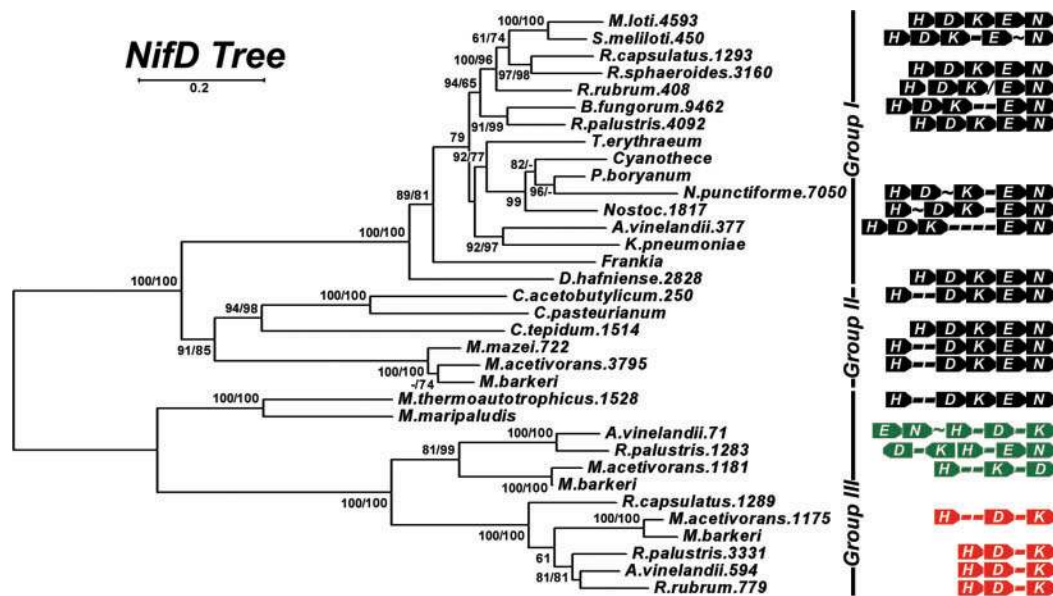
Fig. 4.—NifD phylogeny. Details are as given in figure 3.

after the major lineages diverged, the ensuing nitrogenase/O_2 dilemma would have been difficult to overcome. Despite the abundance of available sequence data from whole genomes, the evolution and distribution of nitrogen fixation in cyanobacteria remains enigmatic and will require more information—genomic, biochemical, and physiological—to unravel. Detailed discussion of this intriguing evolutionary conundrum, and its implications for the role of nitrogen fixatin on the early Earth, can be found especially in Berman-Frank, Lundgren, and Falkowski (2003) as well as in Zehr et al. (2001).

Notably, there are no representatives of diazotrophs among the enteric γ-proteobacteria in this analysis, despite the large number of whole genome sequences from this order that are available. The genus *Klebsiella*, members of whom are frequently equally suited to thrive in terrestrial environments as well as within a host organism, represents the sole best-studied diazotrophic enterobacteriales, and a much anticipated sequencing project (for *K. pneumoniae*) is underway. It is interesting that the emphasis in genome sequencing towards pathogenic microorganisms has resulted in an inadvertent bias against nitrogen fixers (e.g., see fig. 1). Enteric organisms may be utilizing more readily available nitrogenous substrates such as ammonia or glutamate from their host. Under such circumstances, nitrogen fixation genes would be under constant repression and subject to purifying selection and eventually would be lost from the genome of a pathogen that was once diazotrophic. If true, there is no reason to suspect this selective pressure does not exist for all organisms and may account for some of the sporadic taxonomic distribution of nitrogen fixation. Supporting this idea, *Fusobacterium nucleatum*, which is not known to fix nitrogen, has several Nif homologs (HDKE) which, based on sequence analysis, are homologous to Mo-dependent nitrogenases (the NifH protein shows strong similarity, whereas the DKE proteins show weak similarity to known nitrogenase proteins).

However *F. nucleatum* lacks a complete *nif* gene operon and is an oral pathogen, suggesting that it, too, has switched to utilizing less costly pathways for nitrogen assimilation.

Group II

Group II nitrogenases have been well characterized and are very similar in structure and function to their group I homologs (Kim, Woo, and Rees 1993; Leigh 2000). The HDKEN operon structure and synteny is highly conserved, although group II organisms have smaller operons than their group I counterparts. The taxa from this group are distinct from cyanobacteria and proteobacteria in that they are predominantly obligate anaerobes, including methanogens (of the order methanosarciceae), clostridia, and sulfate-reducing bacteria. The monophyly of this cluster is supported especially in NifD primary sequence alignments, where sequences from group II organisms all share an approximately 50-residue conserved insertion, structurally detailed by Kim, Woo, and Rees (1993), that is diagnostic for this group of proteins (Wang, Chen, and Johnson 1988). Gene synteny is also highly conserved among group II, with NifH and NifD consistently separated by two coding genes in clostridial and methanogen operons (see figs. 3–5 and figs. S1and S2 in the Supplementary Material online).

The distinct contrast between group II and 16S-based phylogenetic trees supports the idea that horizontal gene transfer (HGT) has occurred between these organisms. Sequence-based evidence for HGT is corroborated by the ecology of these species; group II organisms are frequently found among syntrophic consortia in anaerobic environments, providing a viable environment for gene sharing (Chien and Zinder 1994; Garcia, Patel, and Ollivier 2000). Protein comparisons between methanogens and clostridia, including and extending beyond nitrogenase, represent several of the most exemplary known cases of HGT
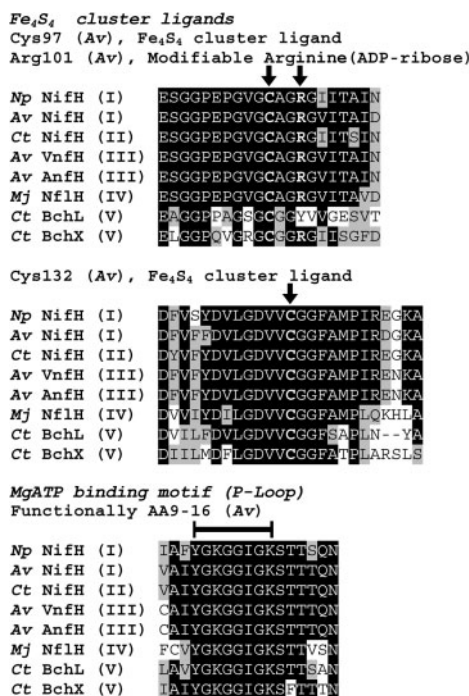
Fig. 5.—NifE phylogeny. Details are as given in figure 3.

(Doolittle 2000). Interestingly, *nif* gene exchange in the archaea seems to have involved only a single order of methanogens (methaniosarcinales) that encompasses *Methanosarcina mazei*, *M. barkeri*, and *M. acetivorans*, although multiple phyla of bacteria have been involved. As discussed below, the phylogenetic resolution of group II and the distinct synapomorphies present (the indels within NifD and NifK) make this an excellent group in which the direction and timing of interdomain horizontal gene transfer can be elucidated.

Group III

The so-called alternative metal, or Mo-independent, nitrogenases, denoted Anf and Vnf, fall within a distinct group III clade. This clade is consistently preserved across all different protein alignments with the sole exception of the VnfH proteins, which are phylogenetically indistinguishable from NifH sequences. In fact, the closest relative to VnfH sequences are typically NifH sequences present in the same genome, as found in *R. palustris* and *A. vinelandii* (fig. 3), indicating relatively recent gene duplication events. It is interesting that the only other distinguishing characteristic used to annotate VnfH sequences are their close association with a vanadium-dependent operon, an occurrence in itself that is not consistent across the organisms examined (e.g., in the *vnf* operon of *R. palustris*). This combined evidence indicates a relatively recent gene duplication origin for VnfH and suggests that there may be some degree of plasticity between NifH and VnfH. This idea is corroborated by recent experimental work where NifH has been effectively replaced by VnfH (Chatterjee et al. 1997).

The remaining group III sequences support an early paralogous origin for alternative nitrogenases. The divergence of Fe-dependent and V-dependent nitrogenases clearly seems to have occurred subsequently, as separate monophyletic clades for these proteins are observed in NifD and NifK trees. One plausible scenario implied by this evidence is that an ancestral NifD homolog might have been less specific with respect to its metal cofactor. Such a nitrogenase would be responsive to environmental

availability of vanadium, iron, and molybdenum that fluctuated with the changing redox state that characterized the Proterozoic Earth between 1 and 2 billion years ago (Normand and Bousquet 1989; Anbar and Knoll 2002). Empirical work supports this idea; under certain conditions FeMoco can be inserted into the Fe-dependent aponitrogenase (which normally requires FeFeco) and vice versa, resulting in a functional, albeit less efficient, enzyme hybrid (Gollan et al. 1993; Pau et al. 1993). Banded iron formations and paleosols indicate that the oceans of the early anoxic Earth were rich in soluble reduced iron, whereas molybdenum, known to be much more soluble in its tetramolybdate (MnO_4^{-2}) form, was likely scarce (Anbar and Knoll 2002). Thus, under the presumably Mo-limiting conditions (or at least Fe-rich conditions) on the early Earth, a nitrogenase able to utilize alternative metals would have been advantageous. As the Precambrian Earth became progressively more oxic because of cyanobacterial photosynthesis, soluble iron became less available and soluble oxidized molybdenum (at least in oxic layers of the ocean) assumed its current role as the most abundant transition metal in the oceans (Anbar and Knoll 2002).

In addition to metal availability, it is interesting that the increased efficiency of Mo-nitrogenase, as compared with the V-dependent and Fe-dependent enzymes, could have served as an additional selection pressure for the Mo-dependent nitrogenase once molybdenum became readily available. This idea addresses one caveat of the concentration scenario; that is, how decreasing soluble iron concentrations characteristic of the middle to late Proterozoic would have been offset by replacing only a single Fe (out of 20 in the NifHDK complex) with molybdenum. A more efficient, Mo-utilizing enzyme could counteract environmental Fe-limitation by achieving the same amount of fixed nitrogen with concomitantly less total nitrogenase synthesized. Hence, modern nitrogenases have likely been refined over hundreds of millions of years of evolution through a combination of these effects, tuned by positive selection both through increasing catalytic efficiency and adapting to changing metal availability. Although here based explicitly upon phylogenetic inference, these ideas on improvements made to nitrogenase through Darwinian

Fe₄S₄ cluster ligands
Cys97 (Av), Fe₄S₄ cluster ligand
Arg101 (Av), Modifiable Arginine(ADP-ribose)

```
              ↓         ↓
Np NifH (I)   ESGGPEPGVGCAGRGIITAIN
Av NifH (I)   ESGGPEPGVGCAGRGVITAID
Ct NifH (II)  ESGGPEPGVGCAGRGIITSIN
Av VnfH (III) ESGGPEPGVGCAGRGVITAIN
Av AnfH (III) ESGGPEPGVGCAGRGVITAIN
Mj NflH (IV)  ESGGPEPGVGCAGRGVITAVD
Ct BchL (V)   EAGGPAGSGCGGYVVGESVT
Ct BchX (V)   ELGGGQVGRGCGGRGIISGFD
```

Cys132 (Av), Fe₄S₄ cluster ligand

```
              ↓
Np NifH (I)   DFVSYDVLGDVVCGGFAMPIREGKA
Av NifH (I)   DFVFDVLGDVVCGGFAMPIRDGKA
Ct NifH (II)  DYVFYDVLGDVVCGGFAMPIREGKA
Av VnfH (III) DFVFYDVLGDVVCGGFAMPIRENKA
Av AnfH (III) DFVFYDVLGDVVCGGFAMPIRENKA
Mj NflH (IV)  DVVIYDILGDVVCGGFAMPLQKHLA
Ct BchL (V)   DVILEDVLGDVVCGGFSAPLN--YA
Ct BchX (V)   DIILMDFLGDVVCGGFATPLARSLS
```

MgATP binding motif (P-Loop)
Functionally AA9-16 (Av)

```
              ├───────┤
Np NifH (I)   IAFYGKGGIGKSTTSQN
Av NifH (I)   VAIYGKGGIGKSTTTQN
Ct NifH (II)  VAIYGKGGIGKSTTTQN
Av VnfH (III) CAIYGKGGIGKSTTTQN
Av AnfH (III) CAIYGKGGIGKSTTTQN
Mj NflH (IV)  FCVYGKGGIGKSTTVSN
Ct BchL (V)   LAVYGKGGIGKSTTSAN
Ct BchX (V)   IAIYGKGGIGKSFTTTN
```

Fig. 6.—Alignment surrounding the MgATP binding motif (bar) and 4Fe-4S coordinating cysteines (vertical arrows) for NifH and NifH homologs from each of the groups in the text (group number after sequence name). Numbering is based on *A. vinelandii* NifH, and black/gray shading indicates 60% or better identity/similarity across all sequences. Np = *N. punctiforme*, Av = *A. vinelandii*, Ct = *C. tepidum*, Mj = *M. jannaschii*.

evolution should be testable through the inference and reconstruction of ancestral nitrogenase sequences. Although the Fe-dependent and V-dependent enzymes are less efficient, organisms such as *R. palustris* and *A. vinelandii* maintain these alternative nitrogenases, likely because of a localized benefit in Fe-rich/Mo-limited conditions, as may be the case in some terrestrial environments (Arial Anbar, personal communication). What remains enigmatic is why all alternative nitrogenases studied so far are found only in organisms that also have Mo-dependent enzymes. In an environment where molybdenum is never available, might one find only alternative nitrogenases? Despite these speculations, what seems clear is that the early evolution of the cognate nitrogenase families has been dramatically affected on multiple levels by the development of oxygenic photosynthesis.

Groups IV and V

As a whole, groups IV and V include a diverse range of NifH and NifD homologs that are not known to be involved with fixing nitrogen. The pigment biosynthesis complexes protochlorophyllide reductase and chlorophyllide reductase, denoted herein as group V nitrogenase homologs, are not only homologous but are functionally analogous to nitrogenase, coupling ATP hydrolysis–driven electron transfer to substrate reduction (Burke, Hearst and Sidow 1993; Fujita and Bauer 2000). As with nitrogenase, electrons flow from a NifH-like ATPase (BchL and BchX)

to a NifDK-like putative heterotetramer where the tetrapyrrole is bound (BchNB and BchYZ). These two enzymes catalyze independent reductions on opposing sides of a tetrapyrrole ring that are essential late steps in chlorophyll and bacteriochlorophyll biosynthesis. Based on sequence similarity, the BchLNB and BchXYZ complexes appear to have originated from a duplicated common ancestor that was less substrate specific and able to catalyze both sequential ring reductions, albeit less efficiently (fig. 2 shows the phylogenetic position of group V pigment biosynthesis genes). Interestingly, both complexes are found together only in anoxygenic photosynthetic bacteria; only one of the two complexes (denoted Bch or Chl LNB in anoxygenic or oxygenic phototrophs, respectively) is found in cyanobacteria, and consequently these organisms produce chlorophyll rather than more reduced bacteriochlorophyll (Xiong et al. 2000; Blankenship 2002). This fortuitous change, likely prompted by loss of the BchXYZ complex in the ancestor of modern cyanobacteria, resulted in a blue shift in primary pigment absorption wavelength, thereby incurring an increase in redox potential of the photosynthetic reaction center. This redox shift provided the necessary energy for the oxidation of water to O_2 via oxygenic photosynthesis, common to cyanobacteria and photosynthetic eukaryotes (Blankenship and Hartman 1998). The subsequent oxidation of the Earth's atmosphere had profound effects on Precambrian life and is widely held to have paved the way for the evolution of complex life (Blankenship and Hartman 1998; Des Marais 2000; Blankenship 2002). Detailed evolutionary trees for pigment biosynthesis genes have been constructed (Xiong et al. 2000) and will not be covered here.

Group IV consists of a subset of nitrogenase homologs (Nif-like proteins, herein designated NflH or NflD, depending on homology) that have yet to be characterized and, based on their breadth in this analysis, appear much more prevalent than previously known. Intriguingly, across the 101 genomes analyzed, these group IV homologs are found only in methanogens—not all of which are diazotrophs—and in some nitrogen-fixing bacteria, most of which are photosynthetic (the sole nonphotosynthetic exception so far being *D. hafniense*, whose nitrogenase proteins are most similar to those from heliobacteria—the only known grampositive phototrophs). Conserved residues in alignments of NifH homologs (fig. 6) from all five groups show that 4Fe-4S iron sulfur cluster-ligating cysteines and the P-loop/MgATP binding motif are invariant, suggesting that these proteins may function analogously to dinitrogenase reductase. Conversely, NifD homologs are highly diverged from both the nitrogenase subunits and the pigment biosynthesis genes. FeMoco-ligating residues (fig. 7) are not conserved among group IV and V proteins, although several—but not all—conserved cysteines involved with P cluster coordination are found in NifD and NifK homologs. This suggests that a less complex FeS cluster, such as a 4Fe-4S, may be functioning in electron transfer in the Group IV and V proteins. In some bacterial genomes, these genes are still found in putative operons, although the operon consists only of a NifH homolog and one or two NifD homologs. However, they are not found in operons among any methanogens, and taken with the unresolved paraphyly
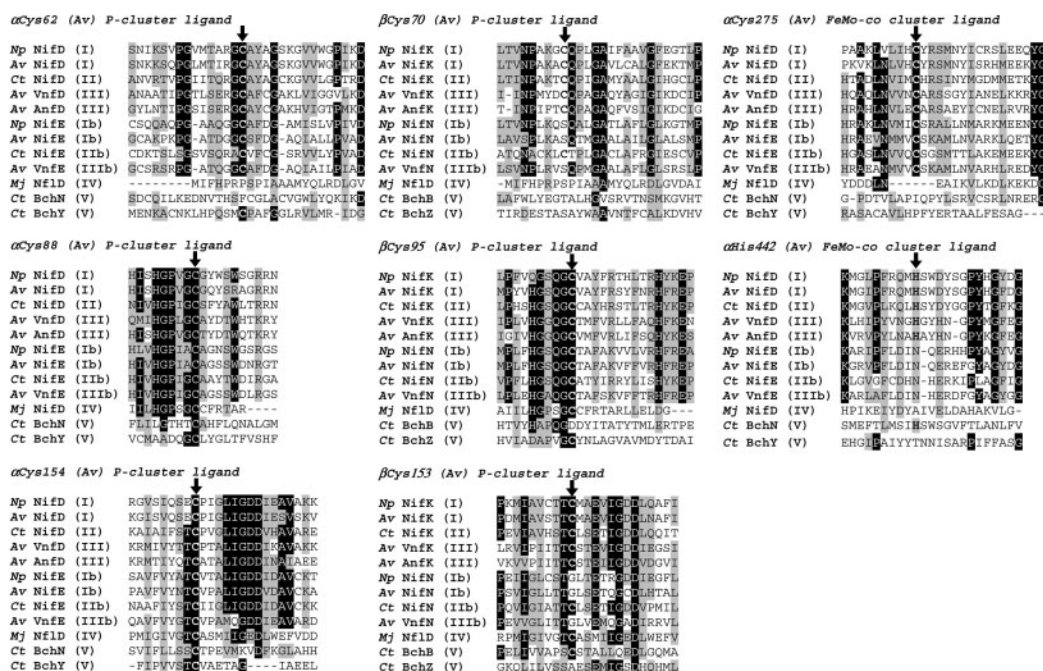
FIG. 7.—Conservation in and around crucial residues (FeMo-co and P-cluster ligands) in NifD (α) and NifK (β) and homologs from each of the five groups (in parenthesis after sequence names) discussed in the text. NifE and NifN sequences are included, respectively, with NifD and NifK alignments (noted with "b" after the group number). The P-cluster and FeMo-co ligands, based on *A vinelandii* numbering, are indicated with vertical arrows. Black/gray shading indicates 60% or better identity/similarity across all sequences. Two = letter binomen abbreviations are as given for figure 6.

and low sequence identities that characterize these phylogenies, it is plausible that these Nif homologs have evolved to different functions in different domains. It is interesting to speculate that this group of proteins, apparently still utilized by a small subset of organisms, may represent the modern vestige of a primitive nitrogenase (a possible intermediary functional link to this enigmatic group of nitrogenase homologs may be provided by *M. barkeri)*, wherein one alternative nitrogenase operon has VnfH, VnfD, and VnfK proteins that cluster with other known vanadium-dependent nitrogenases but whose VnfE and VnfN proteins are only distantly similar to their group III counterparts and, phylogenetically, look like group IV proteins (see Thiel [1996] for characterization and interesting discussion of what may be close homologs of these proteins in cyanobacterium *Anabaena variabilis*).

Several authors have conjectured on the nature of the earliest nitrogenase, suggesting that it may have functioned as nonspecific reductase acting on cyanide, azide, or another N_2 analog (Silver and Postgate 1973; Fani, Gallo and Lio 2000). Such an enzyme might have played a role in carbon and nitrogen assimilation in early microbes, acted as a detoxyase, or alternatively as a redox-sensitive conduit for shuttling excess electrons to a readily available substrate. Ultimately these predictions will be tested through characterization of one or more group IV homologs, currently in progress for *nif* homologs from the non–nitrogen-fixing methanogen *Methanocaldococcus jannaschii* (Staples, Mukhopadhyay, and Blankenship 2003). Understanding the role of these apparently primitive proteins will undoubtedly provide valuable insights into the evolution of nitrogen fixation and photosynthesis, two centrally important metabolic processes on the modern as well as the early Earth.

Plausible Scenarios for Unifying Gene and Species Trees

Fixed nitrogen, particularly ammonia, is essential for life but is thought to have become a limiting nutrient early in the development of life, providing the evolutionary impetus for the development of nitrogenase (see Towe [2002] and reponse). The relative simplicity and phylogeny of group IV nitrogenase homologs suggest that nitrogenase was almost certainly recruited, probably from a multisubunit redox-active protein (but as is typically the case in phylogenetic inference, the reverse is certainly possible). Additionally, the structural and mechanistic similarity between nitrogenase and group V homologs is evidence that gene duplication and recruitment have occurred multiple times in the evolution of this enzyme family. Intriguingly, several of the genomes studied herein, such as that of *R. palustris*, have genes from as many as four of the five groups of nitrogenase homologs. Based on phylogenetic reconstruction as well the presence of nitrogenase in diverse archaea as well as bacteria, it has been inferred that the nitrogenase family had already evolved in the last common ancestor (LCA) of the three domains of life (Normand et al. 1992; Fani, Gallo, and Lio 2000). The LCA hypothesis necessitates that gene loss has been a dominant factor in the extant distribution of nitrogenase, which accounts for the fact that nitrogenase is found neither in eukaryotes nor in many entire phyla of prokaryotes. Importantly, however, the phylogenies inferred here
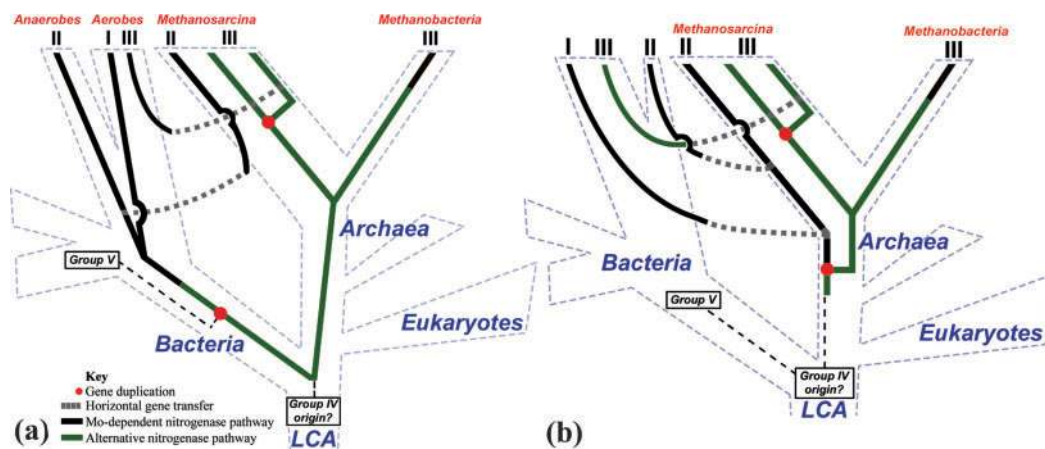
Fig. 8.—Proposed gene versus species trees, as discussed in the text. (*a*) Nitrogen-fixing LCA hypothesis, showing the three domains and their divergence from the LCA in blue text and dashed blue lines. Solid lines (black and green, depending on hypothesized metal specificity) indicate the evolution of nitrogenase from the base of the tree a group IV ancestor. Also indicated are putative gene duplication (red dots) and horizontal gene transfer (originating at gray boxes, with transfer indicated by gray dashed lines) events. The three nitrogenase phylogenetic groups described in the text are indicated at the tips of the tree, as are the predominant organisms in which they are found. According to the LCA model, gene loss has been extensive and accounts for the majority of modern lineages not being able to fix their own nitrogen. (*b*) Methanogen origin hypothesis, using the same color scheme and symbols as figure 8*a*. According to this model, nitrogen fixation was invented in methanogenic archaea and subsequently was transferred into a primitive bacterium, circumventing the necessity for extensive gene loss to explain the paucity of diazotrophic lineages. As with the LCA hypothesis, several relatively recent HGT events must have occurred to explain the distribution and high identities of group II and III nitrogenases.

suggest that horizontal gene transfer has also played a significant role in the history and propagation of nitrogen fixation. The interdomain phylogenetic clustering observed, vis-à-vis group II and group III methanogens and bacteria, is supported by some of the highest identities known between archaeal and bacterial proteins (for example, AnfD proteins are typically more than 75% identical across the two domains). Therefore, the idea that nitrogen fixation had originated in the LCA, at least as inferred by the presence of nitrogenase in the two major prokaryote domains, should be tempered in light of the strong evidence supporting HGT. What can phylogenetic analysis say about this origin and the direction of nitrogenase transfer, and do the data require nitrogen fixation to have been present in the LCA?

The NifH, NifD, and NifE phylogenies shown in figures 3–5 consistently support a clustering of groups I and II, with group I consisting entirely of bacteria and group II consisting of anaerobes and methanogens. Group III also shows evidence of a methanogen/bacterial HGT event, although all phylogenies indicate that it clearly diverged earlier than the group I and II split based on rooting with group IV and V nitrogenase homologs (fig. 2). An evolutionary scheme consistent with the LCA nitrogenase origin is shown in figure 8*a* and described here. As expected if nitrogen fixation were present in the last common ancestor of extant organisms, this solution maps the deep division between groups I/II and group III onto the bacterial/archaeal divergence. Here, gene loss must be invoked as the dominant mechanism accounting for the lack of nitrogenase homologs in eukaryotes and most prokaryotes, putatively spurred on (as discussed above) by the metabolic expense of nitrogen fixation as well as the increasing availablility of abiotically fixed nitrogen once the atmosphere became oxic (Raven and Yin 1998). The subsequent divergence between groups I and II

would have then occurred in the bacteria, perhaps spurred by the development of oxygenic photosynthesis and the resultant aerobic/anaerobic segregation of environments, thereby explaining the aerobe/anaerobe dichotomy of group I versus group II taxa. Subsequently, a group II nitrogenase operon was transferred into a methanosarcina from an anaerobic bacterium, the two species perhaps sharing very similar environmental niches because of metabolically driven symbiosis (discussed below). To our knowledge so far, the group III nitrogenases were maintained only in methanogens, and the primary split within group III appears between methanobacteria and methanosarcina, consistent with speciation (orthologous divergence) between these two families. The emergence of independent Anf and Vnf nitrogenases occurred relatively recently and only in methanosarcina. Subsequent to the Anf/Vnf divergence, both of these alternative nitrogenases were transferred into several different bacterial groups, perhaps multiple times, from a methanosarcina. This transfer marked the first appearance of group III nitrogenase in bacteria, and presumably has allowed organisms to fix nitrogen, albeit less efficiently, in Mo-limited environments.

An interesting aside raised in this scheme is that both HGT events (bacterial group II into methanogens and methanogen group III into several group I bacteria) appear to have carried nitrogenase to organisms that already had a copy of the enzyme. At face value, this undermines the fitness benefit that would be gained by a non–nitrogen-fixing organism acquiring nitrogenase. However, this may be consistent with the regulatory and structural complexity (for example, being able to synthesize and insert intricate metal clusters) inherent to expressing a functional nitrogenase. A diazotrophic organism with these regulatory and assembly components already present (being used for the original nitrogenase copy) would be more likely to utilize

a "new" nitrogenase acquired via HGT. Importantly, these putative HGT events all involve different nitrogenase families, so that either a Mo-dependent nitrogenase is transferred to an organism possessing only Fe-dependent or V-dependent nitrogenases, or vice versa. Thus an organism in an environment with fluctuating trace metal availability might in fact benefit from acquiring a second nitrogenase. Interestingly, under the assumption of an early nitrogenase less specific with regard to its metal dependence, the Mo-dependence seen in methanobacteria (such as *Methanothermobacter thermoautotrophicus* and *Methanococcus maripaludis*) suggests either that molybdenum utilization has occurred independently in multiple lineages or, contrary to paleogeological arguments, that the most primitive nitrogenase was Mo-dependent. On the whole, these ideas suggest that the extraneous (e.g., regulatory and assembly) components of the nitrogenase system will provide an important and informative test bed for understanding the evolutionary history of nitrogenase.

It is worth noting that the root of the three functional nitrogenase groups as illustrated in figure 8*a* could be moved from the LCA up into prokaryotic domains by assuming group III nitrogenases were horizontally transferred very early on into a primitive methanogen. Subsequent bifurcations and gene transfer events would still occur as elucidated above, and the substantial degree of gene loss required by the LCA model is circumvented. Until further evidence can be brought to bear on these earliest events, it is impossible to distinguish a possible bacterial origin from the LCA hypothesis. It may also be that these earliest events cannot be resolved, far buried in the intertwined phylogenetic branches of earliest microbial life that, as some authors have speculated, were both wrought by and fraught with, horizontal gene transfer (Doolittle 1999; Woese 2002). The enigmatic but clear ancestry shared between Mo-dependent group III nitrogenases found in one order of methanogens and alternative nitrogenases found in a second—and the high sequence similarity found among a sporadic distribution of group III bacterial nitrogenases-is extremely important. It supports not only that nitrogen fixation in methanogens was an ancient process—presumably preceding the rise in availability of molybdenum some 2 billion years ago—but also establishes a more parsimonious direction of group III HGT from methanogens into bacteria. Thus, early methanogens had independently evolved both Mo-dependent and Mo-independent nitrogenases, which feasibly took place before any HGT input from early bacteria (the group II HGT event from anaerobes to methanosarcina). Could it be possible that these early methanogen events actually represented the invention of nitrogen fixation, with subsequent HGT carrying nitrogenase to bacteria at a later time? We detail this hypothesis below and find that it also plausibly explains the modern distribution of nitrogenase, without the extensive gene loss necessary in inferring a nitrogen-fixing LCA.

This second scenario hypothesizes that nitrogen fixation per se was invented by methanogenic archaea and subsequently transferred, here in least three separate events, into bacterial lineages. This hypothesis, detailed in figure 8*b*, also would explain the absence of diazotrophs among the eukaryotic domain, crenarchaea, and early branching bacterial lineages not easily accounted for in any LCA-type hypothesis (see fig. 1), and is consistent with the idea that methanogens may have once played a dominant role in the biosphere (discussed below). The divergence of the major groups of methanogens has likely been precipitated in part by increasingly complex metabolic versatility. For example, members of the order methanosarcina (here, *M. acetivorans*, *M. mazei*, and *M. barkeri*) are able to utilize a complex assortment of substrates in methanogenesis, including acetate and a variety of one-carbon compounds, and concomitantly occupy a wide range of environmental niches (Galagan et al. 2002). Conversely, their class methanobacteria cousins, including *M. thermolithotrophicus*, are dependent upon H_2 and CO_2 for methanogenesis and are found in a comparatively narrow range of habitats (Galagan et al. 2002). As illustrated in figure 8*b*, the proposed scheme for a methanogen origin of nitrogen fixation might proceed as follows: At the time when the methanogen families diverged, two paralogous nitrogenase families would have existed, only one of which—the precursor to the group III Anf/Vnf family—was maintained in *M. thermolithotrophicus*. Metal specialization would have occurred later, likely in response to changing trace metal availability (as outlined above). This scenario explicates the basal position of Mo-dependent *nif* genes from *M. thermolithotrophicus* among the group III nitrogenases. The *vnf* and *anf* genes were transferred into bacteria much later, explaining the high identities found between archaeal and bacterial *anf* and *vnf* genes. Additionally (though anecdotal until further supporting evidence is available), *anf* and *vnf* genes in *M. acetivorans* are found in one contiguous operon, so a single archaeal-to-bacterial gene transfer event involving this one large operon could have transferred two nitrogenases to, for example, *R. palustris* and *A. vinelandii*, in which the *anf* and *vnf* operons are not juxtaposed (group III HGT, shown in gray on fig. 8*b*). HGT among the group II phylogeny, promoted by closely shared niches between these organisms, is corroborated by conserved indels in NifD and NifK, gene synteny, and otherwise high sequence identity (although lower than found in Anf/Vnf alignments, suggesting more recent HGT in the latter case). Finally, it is worth noting that all three types of nitrogenase are found in methanogens and, importantly, their sequence evolution and change in metal specificity parallels speciation events within this putatively ancient group of archaea. This parallel evolution of nitrogenase and methanogens, along with the clear evidence for late horizontal gene transfer between methanogens and different and largely unrelated groups of bacteria, lends credence not only to a direction of HGT to a possible origin of nitrogenase within the archaea.

Importantly, the hypotheses on the origin of nitrogen fixation advanced here are only two out of many feasible scenarios, several of which emerge as being equally or nearly as parsimonious as those presented in figure 8*a* and 8*b*. These two scenarios attempt to capture both the phylogenetic and biogeochemical data available and, within these constraints, portray an "LCA/gene loss dominant" versus a "methanogen origin/HGT dominant"

scenario representing the two spectral extremes of many possible evolutionary trajectories.

It is perhaps not surprising that in each of the scenarios outlined above, the methanogens implicated in nitrogenase HGT are methanosarcina, a family demonstrated to be more metabolically versatile and found in comparatively more diverse environments than many of their obligate H_2-utilizing methanobacteria relatives. Lending support to this notion, the genome of methanosarcina *M. mazei* stands as a paradigm for interdomain HGT with roughly a third of its 3.400 ORFs having closest homologs among bacteria (Deppenmeier et al. 2002). Additionally, methanogens are thought to have once played an important if not pivotal role on the early Earth: methanogen-produced methane has been cited as a possible greenhouse gas solution to the faint young sun paradox, as an O_2 sink and as a conduit for H_2 escape into space an impetus for the subsequent oxidation of the atmosphere (Pavlov et al. 2000; Catling, Zahnle, and McKay 2001; Kasting and Siefert 2002). It is feasible that before the Precambrian oxidation event and the so-called "age of cyanobacteria," methanogens may have occupied a dominant niche in the biosphere (Kasting and Siefert 2001). Significantly, at high atmospheric concentrations, methane is known to counter abiotic nitrogen fixation caused by lightning discharge, so the very presence of a growing biomass of methanogens during the Archean could have precipitated a nitrogen crisis, triggering a selection pressure favoring the evolution of nitrogenase (Navarro-Gonzalez, McKay, and Mvondo 2001) (although, importantly, methane can also react with N_2 to form HCN, a viable abiotic source of fixed nitrogen that would potentially counteract the nitrogen crisis [Kasting and Siefert 2001]). If such a crisis did indeed occur, other organisms at the time would have certainly been affected by the dwindling fixed-nitrogen supply and would have to circumvent nitrogen limitations through scavenging (e.g., through association with diazotrophs) or, as some evolutionary evidence points to, by acquiring nitrogenase through horizontal gene transfer. Bacterial-methanogen symbioses, particularly among methanosarcina and aceto-genic bacteria in anoxic environments, have been well studied as important metabolically governed associations and plausibly represent modern analogs of a microbial con-sortium that once dominated the Earth (Kotelnikova 2002). These shared environments, where a diverse consortium of species are juxtaposed within narrow niches, may indeed serve as evolutionary hotspots wherein HGT is signifi-cantly more favorable because of close physical interac-tion between organisms.

Although tenuous because of extensive sequence divergence and concordantly difficult protein alignments, the placement of the root for NifH, NifD, and NifK trees as suggested by groups IV and V nitrogenase homolog outgroups falls consistently between group III and the base of groups I and II. The distribution and phylogeny of (non–nitrogen-fixing) homologs to the HDK nitrogenase complex may suggest that they represent the remnants of a broad family of primitive, prenitrogen fixation proteins. NifE and NifN have an early history consistent with their presence in the LCA, in that group II and group III representatives are more similar to one another than either

is to group I proteins. This contrasts with the HDK phylogeny and may indicate an early function for the NifEN complex, perhaps as a functional nitrogenase transitional step between 4Fe-4S cluster–containing an-cestors of group IV and more complex metal cofactors found in NifDK. Although speculative, this idea explains their subsequent recruitment as scaffolds for FeMoco cluster assembly. Group V represents yet another instance of recruitment of this primitive reductase that likely occurred in the earliest bacterial phototroph and heralded a major step in the evolution of photosynthesis.

Nitrogen fixation is undoubtedly an ancient innova-tion that is not only crucial for extant life, but played a critical role during the early expansion of microbial life as abiotic nitrogen sources became scarce. By considering histories of multiple genes and operons from completely sequenced genomes, it is possible to understand the influence of paralogy, gene recruitment, and horizontal gene transfer in the evolution of nitrogenase. In light of the intractability often thought posed by such complex genetic events, converging lines of biochemical, geological, and phylogenetic evidence make it possible not only to rectify inconsistencies between gene and species trees, but also to elucidate the selective pressures dictating the tempo and mode of organismal versus genomic evolution.

Acknowledgments

Literature Cited

Adams, D. G. 2000. Heterocyst formation in cyanobacteria. Curr. Opin. Microbiol. **3**:618–624.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. **215**:403–410.

Anbar, A. D., and A. H. Knoll. 2002. Proterozoic ocean chem-istry and evolution: a bioinorganic bridge? Science **297**:1137–1142.

Berman-Frank, I., P. Lundgren, Y. B. Chen, H. Kupper, Z. Kolber, B. Bergman, and P. Falkowski. 2001. Segregation of nitrogen fixation and oxygenic photosynthesis in the marine cyanobacterium Trichodesmium. Science **294**:1534–1547.

Berman-Frank, I., P. Lundgren, and P. Falkowski. 2003. Nitrogen fixation and photosynthetic oxygen evolution in cyanobacteria. Res. Microbiol. **154**:157–164.

Blankenship, R. E. 2002. Molecular mechanisms of photosyn-thesis. Blackwell Science, Oxford.

Blankenship, R. E., and H. Hartman. 1998. The origin and evolution of oxygenic photosynthesis. Trends Biochem. Sci. **23**:94–97.

Burke, D. H., J. E. Hearst, and A. Sidow. 1993. Early evolution of photosynthesis: clues from nitrogenase and chlorophyll iron proteins. Proc. Natl. Acad. Sci. USA **90**: 7134–7138.

Catling, D. C., K. J. Zahnle, and C. McKay. 2001. Biogenic methane, hydrogen escape, and the irreversible oxidation of early Earth. Science **293**:839–843.

Chatterjee, R., R. M. Allen, P. W. Ludden, and V. K. Shah. 1997. In vitro synthesis of the iron-molybdenum cofactor and maturation of the nif-encoded apodinitrogenase: fffect of substitution of VNFH for NIFH. J. Biol. Chem. **272**:21604–21608.

Chien, Y. T., and S. H. Zinder. 1994. Cloning, DNA sequencing, and characterization of a nifD-homologous gene from the archaeon Methanosarcina barkeri 227 which resembles nifD1 from the eubacterium Clostridium pasteurianum. J. Bacteriol. **176**:6590–6598.

Cole, J. R., et al. (12 co-authors). 2003. The Ribosomal Database Project (RDP-II): previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy. Nucleic Acids Res. **31**:442–443.

Deppenmeier, U., et al. (22 co-authors). 2002. The genome of *Methanosarcina mazei*: evidence for lateral gene transfer between bacteria and archaea. J. Mol. Microbiol. Biotechnol. **4**:453–61.

Des Marais, D. J. 2000. Evolution. When did photosynthesis emerge on Earth? Science **289**:1703–1705.

Doolittle, R. F. 2000. Searching for the common ancestor. Res. Microbiol. **151**:85–89.

Doolittle, W. F. 1999. Lateral genomics. Trends Cell Biol. **9**: M5–M8.

Egener, T., A. Sarkar, D. E. Martin, and B. Reinhold-Hurek. 2002. Identification of a NifL-like protein in a diazotroph of the beta-subgroup of the Proteobacteria, *Azoarcus* sp. strain BH72. Microbiology **148**:3203–3212.

Falkowski, P. G. 1997. Evolution of the nitrogen cycle and its influence on the biological sequestration of CO_2 in the ocean. Nature **387**:272–275.

Fani, R., R. Gallo, and P. Lio. 2000. Molecular evolution of nitrogen fixation: the evolutionary history of the nifD, nifK, nifE, and nifN genes. J. Mol. Evol. **51**:1–11.

Felsenstein, J. 1989. PHYLIP: Phylogeny inference package. Version 3.2. Cladistics **5**:164–166.

Fujita, Y., and C. E. Bauer. 2000. Reconstitution of light-independent protochlorophyllide reductase from purified bchl and BchN-BchB subunits. In vitro confirmation of nitrogenase-like features of a bacteriochlorophyll biosynthesis enzyme. J. Biol. Chem. **275**:23583–23588.

Galagan, J. E., et al. (55 co-authors). 2002. The genome of *M. acetivorans* reveals extensive metabolic and physiological diversity. Genome Res. **12**:532–542.

Garcia, J. L., B. K. C. Patel, and B. Ollivier. 2000. Taxonomic phylogenetic and ecological diversity of methanogenic Archaea. Anaerobe **6**:205–226.

Gollan, U., K. Schneider, A. Muller, K. Schuddekopf, and W. Klipp. 1993. Detection of the in vivo incorporation of a metal cluster into a protein. The FeMo cofactor is inserted into the FeFe protein of the alternative nitrogenase of *Rhodobacter capsulatus*. Eur. J. Biochem. **215**:25–35.

Goodman, R. M., and J. B. Weisz. 2002. Plant-Microbe Symbioses: An Evolutionary Survey. Biodiversity of Microbial Life: Foundation of Earth's Biosphere. J. T. Staley and A. L. Reysenbach, eds. Wiley, New York.

Halbleib, C. M., and P. W. Ludden. 2000. Regulation of biological nitrogen fixation. J. Nutr. **130**:1081–1084.

Hirsch, A. M., H. I. McKhann, A. Reddy, J. Liao, Y. Fang, and C. R. Marshall. 1995. Assessing horizontal transfer of nifHDK genes in eubacteria: nucleotide sequence of nifK from Frankia strain HFPCcI3. Mol. Biol. Evol. **12**:16–27.

Imhoff, J. F. 1995. Taxonomy and physiology of photosynthetic purple bacteria and green sulfur bacteria. Pp. 1–15 *in* R. E. Blankenship, M. T. Madigan and C. E. Bauer, eds. Anoxygenic photosynthetic bacteria. Kluwer, Dordrecht, the Netherlands.

Joerger, R. D., and P. E. Bishop. 1988. Bacterial alternative nitrogen fixation systems. Crit. Rev. Microbiol. **16**:1–14.

Kasting, J. F., and J. L. Siefert. 2001. Biogeochemistry: the nitrogen fix. Nature **412**:26–27.

———. 2002. Life and the evolution of Earth's atmosphere. Science **296**:1066–1068.

Kessler, P. S., C. Blank, and J. A. Leigh. 1998. The *nif* gene operon of the methanogenic archaeon *Methanococcus maripaludis*. J. Bacteriol. **180**:1504–1511.

Kessler, P. S., and J. A. Leigh. 1999. Genetics of nitrogen regulation in *Methanococcus maripaludis*. Genetics **152**: 1343–1351.

Kim, J., D. Woo, and D. C. Rees. 1993. X-ray crystal structure of the nitrogenase molybdenum-iron protein from *Clostridium pasteurianum* at 3.0-A resolution. Biochemistry **32**:7104–7115.

Kotelnikova, S. 2002. Microbial production and oxidation of methane in deep subsurface. Earth Sci. Rev. **58**:367–395.

Kumar, S., K. Tamura, and M. Nei. 1994. MEGA: molecular evolutionary genetics analysis software for microcomputers. Comput. Appl. Biosci. **10**:189–191.

Leigh, J. A. 2000. Nitrogen fixation in methanogens: the archaeal perspective. Curr. Issues Mol. Biol. **2**:125–31.

Miller, R. W., and R. R. Eady. 1988. Molybdenum and vanadium nitrogenases of *Azotobacter chroococcum*. Low temperature favours N2 reduction by vanadium nitrogenase. Biochem. J. **256**:429–432.

Navarro-Gonzalez, R., C. P. McKay, and D. N. Mvondo. 2001. A possible nitrogen crisis for Archaean life due to reduced nitrogen fixation by lightning. Nature **412**:61–64.

Normand, P., and J. Bousquet. 1989. Phylogeny of nitrogenase sequences in Frankia and other nitrogen-fixing microorganisms. J. Mol. Evol. **29**:436–447.

Normand, P., M. Gouy, B. Cournoyer, and P. Simonet. 1992. Nucleotide sequence of nifD from *Frankia alni* strain ArI3: phylogenetic inferences. Mol. Biol. Evol. **9**:495–506.

Pau, R. N., M. E. Eldridge, D. J. Lowe, L. A. Mitchenall, and R. R. Eady. 1993. Molybdenum-independent nitrogenases of *Azotobacter vinelandii*: a functional species of alternative nitrogenase-3 isolated from a molybdenum-tolerant strain contains an iron-molybdenum cofactor. Biochem. J. **293**(pt 1): 101–107.

Pavlov, A. A., J. F. Kasting, L. L. Brown, K. A. Rages, and R. Freedman. 2000. Greenhouse warming by CH4 in the atmosphere of early Earth. J. Geophys. Res. **105**:11981–11990.

Prakash, R. K., R. A. Schilperoort, and M. P. Nuti. 1981. Large plasmids of fast-growing rhizobia: homology studies and location of structural nitrogen fixation (nif) genes. J. Bacteriol. **145**:1129–1136.

Raven, J. A., and Z.-H. Yin. 1998. The past, present and future of nitrogenous compounds in the atmosphere, and their interactions with plants. New Phytol. **139**:205–219.

Roll, J. T., V. K. Shah, D. R. Dean, and G. P. Roberts. 1995. Characteristics of NIFNE in *Azotobacter vinelandii* strains: implications for the synthesis of the iron-molybdenum cofactor of dinitrogenase. J. Biol. Chem. **270**:4432–4437.

Schmidt, H. A., K. Strimmer, M. Vingron, and A. von Haeseler. 2002. TREE-PUZZLE: maximum likelihood phylogenetic

analysis using quartets and parallel computing. Bioinformatics **18**:502–504.

Silver, W. S., and J. R. Postgate. 1973. Evolution of asymbiotic nitrogen fixation. J. Theor. Biol. **40**:1–10.

Simpson, F. B., and R. H. Burris. 1984. A nitrogen pressure of 50 atmospheres does not prevent evolution of hydrogen by nitrogenase. Science **224**:1095–1097.

Staples, C., B. Mukhopadhyay, and R. E. Blankenship. 2003. The evolutionary relationship between nitrogen fixation and bacteriochlorophyll biosynthesis. 2003 General Meeting of the NASA Astrobiology Institute, Tempe, AZ. Astrobiology (abstract) **2**(4):496–497.

Thiel, T. 1993. Characterization of genes for an alternative nitrogenase in the cyanobacterium *Anabaena variabilis*. J. Bacteriol. **175**:6276–6286.

———. 1996. Isolation and characterization of the VnfEN genes of the cyanobacterium *Anabaena variabilis*. J. Bacteriol. **178**:4493–4499.

Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22**:4673–4680.

Towe, K. M. 2002. Evolution of nitrogen fixation. Science **295**:798–799.

Wang, S. Z., J. S. Chen, and J. L. Johnson. 1988. Distinct structural features of the alpha and beta subunits of nitrogenase molybdenum-iron protein of *Clostridium pasteurianum*: an analysis of amino acid sequences. Biochemistry **27**:2800–2810.

Woese, C. R. 2002. On the evolution of cells. Proc. Natl. Acad. Sci. USA **99**:8742–8747.

Xiong, J., W. M. Fischer, K. Inoue, M. Nakahara, and C. E. Bauer. 2000. Molecular evidence for the early evolution of photosynthesis. Science **289**:1724–1730.

Zehr, J. P., M. T. Mellon, and W. D. Hiorns. 1997. Phylogeny of cyanobacterial nifH genes: evolutionary implications and potential applications to natural assemblages. Microbiology **143**:1443–1450.

Zehr, J. P., J. B. Waterbury, P. J. Turner, J. P. Montoya, E. Omoregie, G. F. Steward, A. Hansen, and D. M. Karl. 2001. Unicellular cyanobacteria fix N2 in the subtropical north Pacific Ocean. Nature **412**:635–638.

Zehr, J. P., M. Wyman, V. Miller, L. Duguay, and D. G. Capone. 1993. Modification of the Fe protein of nitrogenase in natural-populations of Trichodesmium-Thiebautii. Appl. Environ. Microbiol. **59**:669–676.