Electron Transfer to Nitrogenase in Different Genomic and Metabolic Backgrounds


ABSTRACT

Nitrogenase catalyzes the reduction of dinitrogen (N₂) using low-potential electrons from ferredoxin (Fd) or flavodoxin (Fld) through an ATP-dependent process. Since its emergence in an anaerobic chemolithotroph, this oxygen (O₂)-sensitive enzyme complex has evolved to operate in a variety of genomic and metabolic backgrounds, including those of aerobes, anaerobes, chemotrophs, and phototrophs. However, whether pathways of electron delivery to nitrogenase are influenced by these different metabolic backgrounds is not well understood. Here, we report the distribution of homologs of Fds, Flds, and Fd-/Fld-reducing enzymes in 359 genomes of putative N₂ fixers (diazotrophs). Six distinct lineages of nitrogenase were identified, and their distributions largely corresponded to differences in the host cells’ ability to integrate O₂ or light into energy metabolism. The predicted pathways of electron transfer to nitrogenase in aerobes, facultative anaerobes, and phototrophs varied from those in anaerobes at the levels of Fds/Flds used to reduce nitrogenase, the enzymes that generate reduced Fds/Flds, and the putative substrates of these enzymes. Proteins that putatively reduce Fd with hydrogen or pyruvate were enriched in anaerobes, while those that reduce Fd with NADH/NADPH were enriched in aerobes, facultative anaerobes, and anoxygenic phototrophs. The energy metabolism of aerobic, facultatively anaerobic, and anoxygenic phototrophic diazotrophs often yields reduced NADH/NADPH that is not sufficiently reduced to drive N₂ reduction. At least two mechanisms have been acquired by these taxa to overcome this limitation and to generate electrons with potentials capable of reducing Fd. These include the bifurcation of electrons or the coupling of Fd reduction to reverse ion translocation.

IMPORTANCE

Nitrogen fixation supplies fixed nitrogen to cells from a variety of genomic and metabolic backgrounds, including those of aerobes, facultative anaerobes, chemotrophs, and phototrophs. Here, using informatics approaches applied to genomic data, we show that pathways of electron transfer to nitrogenase in metabolically diverse diazotrophic taxa have diversified primarily in response to host cells’ acquired ability to integrate O₂ or light into their energy metabolism. The acquisition of two key enzyme complexes enabled aerobic and facultatively anaerobic phototrophic taxa to generate electrons of sufficiently low potential to reduce nitrogenase: the bifurcation of electrons via the Fix complex or the coupling of Fd reduction to reverse ion translocation via the Rhodobacter nitrogen fixation (Rnf) complex.

KEYWORDS

fix, nitrogenase, Rnf, ferredoxin, flavodoxin, hydrogen, nitrogen fixation, oxygen, photosynthesis, pyruvate, bifurcation
Nitrogenase is an oxygen-sensitive enzyme that catalyzes the reduction of dinitrogen (N₂) to ammonia (NH₃), accounting for nearly two-thirds of fixed nitrogen (N) on earth today (1, 2) and thereby modulating the global fixed N supply (3, 4). Nitrogenase is comprised of two components, a homodimeric iron protein (H subunit) and a heterotetrameric dinitrogenase reductase protein complex (D and K subunits) (5, 6). Homologs of nitrogenase are widely distributed among Bacteria but have only been identified in one group of taxa within the archaeal Euryarchaeota phylum, methanogens. Nitrogenase homologs have not been identified among eukarya (1, 7–12). Three types of nitrogenase have been described that are evolutionarily and structurally related but differ in the metallic composition of their active site: molybdenum (Mo)-containing nitrogenases (Nif), vanadium (V)-containing nitrogenases (Vnf), and iron (Fe)-containing nitrogenases (Anf) (13–16).

Phylogenetic evidence indicates that Nif emerged prior to Anf and Vnf and that the earliest-evolving Nif lineages are from anaerobic, hyperthermophilic methanogens, implicating a chemotrophic origin for N₂ fixation in an anoxic environment (7, 9, 13, 16, 17). N₂-fixing microorganisms (i.e., diazotrophs) have since diversified to function in a wide variety of genomic and metabolic backgrounds, including those of aerobes, facultative anaerobes, and phototrophs. Given the O₂ sensitivity of Nif, aerobic, facultatively anaerobic, and oxygenic phototrophic diazotrophs have had to evolve or acquire one of several mechanisms to mitigate the toxicity of O₂ to this enzyme during N₂ fixation (17). These mechanisms include fixing N₂ only during dark hours when oxygenic photosynthesis ceases and when heterotrophic respiration keeps O₂ tensions low (18), using specialized cells called heterocysts in filamentous Cyanobacteria to spatially localize nitrogenase in anaerobic cellular compartments (19), maintaining an anoxic environment in the cytoplasm through increased O₂-dependent respiratory activity (20), or fixing N₂ only during anaerobic growth. A recent study showed that the diversification of Nif in aerobes and facultative anaerobes was also accompanied by the recruitment and loss of genes that are involved in regulating and protecting Nif against oxidative stress (17).

The primary electron donors to Nif are reduced ferredoxin (Fd) (21–24) and flavodoxin (Fld) (25–27). The transfer of eight electrons from Fd or Fld to NifH and ultimately to NifDK is an ATP-dependent process, requiring at a minimum 16 mol ATP per mol N₂ reduced (28–30). Less is known of the stoichiometry of ATP hydrolysis per mol N₂ reduced by Anf or Vnf (31–34). The reduction of Fd or Fld in diazotrophs that occupy anoxic environments (e.g., clostridia and methanogens) occurs via the activity of pyruvate-flavodoxin oxidoreductase (PFOR) (35–38), a select subset (groups 3c, 3d, 4d, and 4e) of [NiFe]-hydrogenases, or [FeFe]-hydrogenases (39–45). However, the reduction of Fd (standard electrode potential [Eₒ] = −420 mV) in aerobic and some anoxygenic phototrophic diazotrophs is more of a challenge. Rather than producing reduced Fd during their primary energy metabolism, aerobic bacteria and some anoxygenic phototrophs, in particular facultatively anaerobic purple sulfur and nonsulfur bacteria, generate reduced NADH or NADPH (Eₒ = −320 mV) (46, 47), which are not of low enough potential to drive N₂ reduction (48–50). Anaerobic purple sulfur and facultatively anaerobic nonsulfur anoxygenic phototrophic bacteria utilize photosystems to drive cyclic electron transfer. The reduction of NAD⁺/NADH is typically accomplished with electrons supplied by the oxidation of an inorganic substrate (e.g., sulfide, thiosulfate, or hydrogen [H₂]) or by the oxidation of organic compounds and energy from reverse electron transport if the inorganic electron donor is not of low-enough potential to reduce NAD⁺/NADH (51, 52). It has been proposed that these taxa acquired the Fix complex (encoded by fixABCX) and/or the Rhodobacter nitrogen fixation [Rnf, encoded by rnfABCDEG(H)] complex in order to generate reduced Fd from NADH/NADPH (46). Fix catalyzes the oxidation of two NADH to generate a reduced quinone, which fuels the respiratory chain, and a reduced Fd (27, 53, 54), whereas Rnf catalyzes the NADH-dependent reduction of Fd by coupling it to the depletion of the electrochemical gradient (55–58). Oxygenic phototrophic Cyanobacteria utilize photosystem I to energize electrons to
potentials negative enough to drive the reduction of Fd (51, 52). However, this Fd is not available to Nif, since it must be temporally or spatially separated from oxygenic photosynthesis due to inhibition of Nif by O₂. Rather than coding for [FeFe]-hydrogenase, [NiFe]-hydrogenase, Fix, or Rnf and using these enzymes to reduce Fd, *Cyanobacteria* encode ferredoxin-NADP⁺ oxidoreductase (FNR), which can function in reverse to reduce Fd or Fld with NADPH generated by carbohydrate oxidation in heterocysts or when O₂ tensions are low (35, 59–62). Some *Cyanobacteria* also encode PFOR, which might be expected to contribute to Fd reduction in these cells, and this might be used by Nif. Like *Cyanobacteria* anaerobic anoxygenic green sulfur bacteria use a type I photosystem that is distantly related to photosystem I to generate reduced Fd as a component of photosystem-driven cyclic electron transfer (51, 52); however, unlike *Cyanobacteria* it is possible that this serves as a reductant for N₂ fixation (63). Green sulfur bacteria also encode an FNR, which is phylogenetically and structurally unrelated to conventional FNR found in *Cyanobacteria* and this can also be used to drive reduction of Fd (64, 65). Thus, at least seven enzyme complexes have evolved to provide reduced Fd for N₂ fixation: PFOR, [NiFe]-hydrogenase, [FeFe]-hydrogenase, Rnf, Fix, and both forms of FNR (35); however, their distribution in the genomes of diazotrophs is not known.

Different Fds and Flds are also likely to be involved in the delivery of electrons to nitrogenase in aerobes, anaerobes, and phototrophs, and these Fds and Flds may vary alongside the primary enzymes that are involved in reducing these electron carriers. Fds are sensitive to O₂ due to the lability of their iron sulfur (FeS) clusters (66–69). In contrast, Flds contain flavin mononucleotide as the prosthetic group involved in electron transfers instead of FeS clusters and, hence, are thought to be less sensitive to O₂ (67, 70, 71). Previous bioinformatics analyses have shown that NifF (17), an Fld that can donate electrons to Nif (24, 72, 73), was recruited to *nif* operons during the diversification of Nif from anaerobic to aerobic taxa (17), which may point to the use of Flds as an adaptive strategy to fix N₂ in oxic environments. Moreover, under the iron-deficient conditions that characterize most circumneutral oxic environments, diazotrophs tend to synthesize Flds preferentially as primary electron donors to Nif (22–24, 26, 74). Studies have also shown that electron delivery by an Fd or Fld can be complemented by other Fds or Flds that are encoded in the genomes of diazotrophs (24, 35, 74–76). These observations suggest that pathways that mediate electron flow to nitrogenase are flexible and vary according to the genomic and metabolic background of taxa.

To better define the systems of electron transfer to nitrogenase in diverse microbes, we compiled all Fd and Fld homologs in Nif-encoding genomes and classified them using homology-based methods. In addition, we compiled homologs of all enzymes that have been shown to reduce Fd or Fld for use in the reduction of N₂ by nitrogenase. These include PFOR, [NiFe]-hydrogenase, [FeFe]-hydrogenase, Fix, Rnf, and both forms of FNR. Since Nif is encoded in all genomes that encode Anf and Vnf (1), we also examined the systems of electron transfer to these alternative nitrogenase isoforms. Organisms encoding Nif, Anf, and Vnf were classified phylogenetically and physiologically as aerobes, anaerobes, or facultative anaerobes and as chemotrophs, anoxic phototrophs, or oxygenic phototrophs based on published data. Statistical analyses were then applied to this curated data set to identify patterns of cooccurrence between the distribution of nitrogenase lineages/isoforms, enzymes that putatively reduce Fd/Fld, and Fds/Flds. The results are discussed in the context of the metabolism of the cells, specifically the influence of O₂ and light on putative pathways of electron delivery to nitrogenase.

**RESULTS AND DISCUSSION**

*Taxonomic distribution and phylogeny of Nif/Anf/Vnf HDK homologs.* Nitrogenases were identified in the genomes of diverse *Bacteria* and *Archaea* that included obligate aerobes, facultative anaerobes, obligate anaerobes, phototrophs and chemotrophs, and autotrophs and heterotrophs. The identification of Nif in organisms with
diverse metabolisms is consistent with a complex evolutionary history that has been described previously for N2 fixation (see Table S1A in the supplemental material) (1, 7, 8, 13, 16). Of the total 4,588 publicly available genomes in our database, 359 genomes (7.8% of the total) encoded the minimal set of proteins for nitrogen fixation (i.e., homologs of NifHDKENB) (7). Forty-six of the 359 taxa with genomes that encode nitrogenase complements have been experimentally shown to grow with atmospheric N2 as their sole N source (Table S1B). Of these 359 nitrogenase-encoding genomes, 48 belonged to obligately chemotrophic and anaerobic Archaea, all of which were affiliated with methanogens within the phylum Euryarchaeota. The remaining 311 diazotrophic genomes were identified in the bacterial domain, with the majority identified as members of the Proteobacteria (n = 191) and Firmicutes (n = 68) (Table S1A). Of the 311 diazotrophic bacterial genomes, 31% were from aerobes, 40% were from facultative anaerobes, and 28% were from obligate anaerobes. Furthermore, 79% of the diazotrophic bacterial genomes were from chemotrophs and 21% were from phototrophs (both oxygenic and anoxygenic).

Phylogenetic reconstruction of a concatenation of HDK proteins revealed six distinct lineages of nitrogenases (Fig. 1). These included the four lineages that have been identified in previous studies (1, 8, 13, 14, 16, 77), which include two Nif sublineages (designated Nif-A and Nif-B) and two sublineages designated Anf and Vnf. In addition, previous combined informatics and structural analyses have suggested that two biochemically uncharacterized lineages likely harbor a molybdenum cofactor (9), and organisms that encode them have experimentally been shown to fix N2 (11, 78); hence, we refer to these lineages as Nif-C and Nif-D in this study.

A total of 224 genomes encoded Nif-A, and these were from taxa that are primarily from the Proteobacteria (n = 169), Firmicutes (n = 26), and Cyanobacteria (n = 20) (Table S1A). Thirty of these taxa have been experimentally shown to fix N2 (Table S1B). However, the cyanobacterium Microcoleus chthonoplastes (which is not included in our database since a complete genome is not available) does encode Nif-A and the minimal set of proteins that allow for N2 fixation (i.e., homologs of NifHDKENB) (7). And yet, cultivation studies suggest that this taxon cannot grow with N2 as its sole N source (79), indicating that the presence of a full nif gene complement does not guarantee the ability to fix N2. The majority of the taxa that encode Nif-A are aerobes (42%) or facultative anaerobes (38% of total taxa) (Fig. 2A). In addition, 23% of Nif-A-encoding genomes were from phototrophs and 77% were from chemotrophs (Fig. S1). Among the 23% of Nif-A homologs that are encoded in the genomes of phototrophs, 9% are
from Cyanobacteria, 11% are from facultatively anaerobic anoxygenic purple nonsulfur bacteria, and 2% are from anaerobic anoxygenic purple bacteria. However, it is not clear how rigorously O₂ usage has been characterized in these anaerobic anoxygenic purple bacteria, which include the following strains: Allochromatium vinosum, Thiocystis violascens, Thioflavicoccus mobilis, and Halorhodospira halophila.

The majority of the 106 genomes that encode Nif-B were from obligate anaerobes (93% of total taxa) and chemotrophs (87% of the total taxa) and anaerobic anoxygenic green phototrophic bacteria (11% of total). Nif-B-encoding taxa were primarily affiliated with Firmicutes (n = 34), Euryarchaeota (n = 29), and Proteobacteria (n = 22). Of these 106 Nif-B-encoding genomes, 13 were from taxa (primarily from the genus Clostridium) that have been experimentally shown to fix N₂ (Table S1B). All of the anaerobic anoxygenic phototrophs were affiliated with the Chlorobia or Chloroflexi (n = 12). Members of the Firmicutes (n = 8) and Euryarchaeota (n = 6) encoded Nif-C (n = 15), while all (n = 13) Nif-D-encoding genomes were methanogens (Euryarchaeota). None of the taxa that encode Nif-C have been shown to fix N₂, while three organisms that encode Nif-D (all of which are methanogens) have been experimentally shown to fix N₂ (Table S1B). Taxa that encode Nif-C were primarily anaerobes (88% of the total taxa), while taxa that encode Nif-D were all anaerobes. All the Nif-C- and Nif-D-encoding genomes were chemotrophs.

A total of 32 genomes encoded Anf, and these were primarily from the Proteobacteria (n = 16) and Firmicutes (n = 11), while a total of 23 genomes encoded Vnf. Most of the Vnf-encoding genomes were from the Euryarchaeota (n = 11) and Firmicutes (n = 8). A separate lineage comprising HDK proteins from taxa that have not yet been shown to fix N₂ were nested among Nif sublineages. These proteins were termed “uncharacterized,” as previously described (13), and were affiliated with members of the Chloroflexi (i.e., Roseiflexus spp.).

Multiple forms of nitrogenase were often detected in the same genome (Fig. 2B). However, the genomes that coded for Nif-C or Nif-D did not code for any other forms of nitrogenases. The genome of the cyanobacterium Pleurocapsa sp. strain PCC 7327 was found to encode both Nif-A and Nif-B. All Anf- and Vnf-encoding genomes also encoded Nif, which is consistent with previous observations (1, 7, 13). Four genomes were identified that coded for Anf, Vnf, and Nif-B, while two genomes encoded Anf, Vnf, and Nif-A. Interestingly, Anf was more commonly detected in Nif-A-encoding genomes (65% of total Anf-encoding genomes), while Vnf was more commonly detected in Nif-B-encoding genomes (77% of total Vnf-encoding genomes). If the root of the nitrogenase phylogeny is within the Nif-D lineage, as has been suggested previously (7, 13, 16), our phylogenetic reconstruction indicates that Anf diverged from a Vnf-like ancestor, both of which emerged from a Nif-C- or Nif-D-like ancestor. Likewise, our phylogenetic reconstruction indicates that Nif-A diversified from a Nif-B-like ancestor.
Taxonomic distribution of ferredoxin and flavodoxin homologs in the genomes of putative diazotrophs. (i) Ferredoxin. The taxonomic distribution of all 47 Fds identified in the genomes of putative diazotrophs is detailed in Table S2 in the supplemental material. A total of 36 Fds were detected in Nif-A-encoding genomes, of which FdxB and FdxA were the most common; they were present in >40% of those genomes (Fig. S2A). fdxB is encoded near nifQ and nifB in the genome of Azotobacter vinelandii (80, 81) and is expressed at a similar level to these genes under N2-fixing conditions (82). Mutational studies have shown involvement of FdxB in active-site metallocluster biosynthesis in A. vinelandii (83, 84). Furthermore, FdxB was shown to be incapable of serving as an electron donor to nitrogenase in vitro using protein from Rhodobacter capsulatus (83). In contrast, in the cyanobacterium Anabaena sp. strain PCC 7120, FdxB has been shown to complement fdxH mutant strains; FdxH is the preferred donor of electrons to nitrogenase in this taxon (85). These observations suggest that FdxB has multiple roles in diazotrophic cells. The prevalence of FdxA in Nif-A-encoding genomes is consistent with its colocalization near major nif gene clusters, such as in the case of A. vinelandii (24), or within the nif gene cluster itself, such as in Herbaspirillum seropedicae (86). This agrees with findings from multiple studies that have documented the ability of FdxA to donate electrons to nitrogenase (87–90). Several less commonly detected Fds (identified in >40% of the Nif-A-encoding genomes) have also been shown or predicted to be involved in electron transfer to nitrogenase. These include FdxN in Rhodospirillum rubrum and Rhizobium mellioti (91, 92), FdxE in Rhodobacter capsulatus (93), FdxH (85, 94), FdxI in Anabaena sp. PCC 7120 (95, 96), and FdxC in R. capsulatus (Fig. S2A) (97). Together, these results suggest that FdxA and FdxB may have a role in electron delivery to Nif-A-like nitrogenases; however, other Fds could potentially complement the functionality of these Fds.

Two Fds found in Clostridium pasteurianum, CpFd1 (65% of the total Nif-B-encoding genomes) and CpFd4 (48% of the total genomes), correlated with the distribution of Nif-B in diazotrophic genomes (Fig. S2A), suggesting a role for these Fds in electron delivery to Nif-B-like nitrogenases. FdxA (47% of the total Nif-C-encoding genomes) and three Fds found in Caldicellulosiruptor saccharolyticus, CsFd3 (47% of the total genomes), CsFd6 (47% of the total genomes), and CsFd1 (41% of the total genomes), were predominant in Nif-C-encoding genomes, while Nif-D-encoding genomes encoded multiple Fds found in Methanocaldococcus vulcanius (MvFds) (Fig. S2A). The distribution of Anf in genomes was only moderately correlated with the distribution of Fds, with FdxA and FdxB yielding the highest correlations (i.e., >45% of the genomes) (Fig. S2A). Finally, the distribution of Vnf was highly correlated with the distribution of CpFd4 (80% of the total genomes) and, to lesser extents, four Fds found in Methanosarcina barkeri, MbFd1, MbFd2, MbFd3, and MbFd4 (Fig. S2A). While electron transfer to Nif-A-like nitrogenases has been studied extensively, we are unaware of experimental data on the role of Fds in electron transfer to Nif-B, Nif-C, Nif-D, Anf, or Vnf, which precludes a comparison of our informatics-based predictions to experimental data.

The Fds identified in our study varied markedly among genomes coding for the various isoforms of nitrogenase (Fig. S2A). Importantly, genomes encoding Nif-A nitrogenases coded for unique Fds that were either absent or present but rarely detected in the genomes of other diazotrophs (Fig. S2A). For example, of 14 abundant Fds (i.e., present in >40% of the genomes) encoded by Nif-A-encoding genomes, 12 were unique to Nif-A-encoding genomes, while only two were identified in the genomes of other diazotrophs (Fig. S2A). In contrast, considerable overlap was observed in the composition of Fds/Flds encoded in genomes that also encoded Nif-B, Nif-C, Nif-D, Anf, or Vnf, which precludes a comparison of our informatics-based predictions to experimental data.

(ii) Flavodoxin. We identified a total of five phylogenetically distinct Flds among nitrogenase-encoding genomes. The taxonomic distribution of all five Flds identified is detailed in Table S2 in the supplemental material. Of the five Flds detected in Nif-A-encoding genomes,FldA was the most common and was identified among 26% of the genomes (Fig. S2B). The distribution of CpFld2 and CpFld3 cooccurred with Nif-B in >40% of the genomes, while CpFld3 was detected in >40% of the Nif-C- and Nif-D-
encoding genomes (Fig. S2B). Like Nif-A-encoding genomes, FldA was the dominant Fld in Anf-encoding genomes (Fig. S2B). Lastly, CpFld3 and CpFld2 were the dominant Flds in Vnf-encoding genomes (Fig. S2B). There is no experimental evidence that these Flds are involved in electron transfer to nitrogenase, with the exception of NifF (present in a few Nif-A-encoding genomes), which has been shown to transfer electrons to nitrogenase in *Klebsiella pneumoniae* (98, 99) and *A. vinelandii* (24, 27).

The distribution of the five Flds identified in this study varied in diazotrophs with different metabolic backgrounds. Of the five Flds identified among the genomes of diazotrophs, NifF and FldA were only detected in the genomes of aerobes, facultative anaerobes, or phototrophs that encoded Nif-A (Fig. S2B). The sole exception to this observation was in the genome of the anaerobic spirochete *Spirochaeta smaragdinae* strain DSM 11293, which was found to code for FldA and Nif-B (Table S2). The remaining three Flds were common in the genomes of anaerobic chemotrophs that encoded Nif-B, Nif-C, and Nif-D. This observation is like that made for Fds, where the distribution of Flds in genomes that encode Nif-A were distinct from those that encoded Nif-B, Nif-D, and Nif-D. These collective observations indicate that the Fds/Flds in aerobic/facultatively anaerobic diazotrophs differ from those in anaerobic diazotrophs, leading to the hypothesis that O2 played a role in the diversification of systems for the delivery of electrons to nitrogenase. Moreover, the Fds/Flds are different among phototrophs and chemotrophs, leading to the hypothesis that the integration of light into the energy metabolism of diazotrophs played a key role in the diversification of systems of electron delivery to nitrogenase.

**Distribution of ferredoxin- and flavodoxin-reducing protein homologs in the genomes of putative diazotrophs.** The taxonomic distributions of PFOR, [NiFe]-hydrogenase, [FeFe]-hydrogenase, Rnf, Fix, and FNR homologs are detailed in Table S2. Briefly, genomes that encoded Nif-A and Nif-B coded for homologs of six of the seven putative Fd- or Fld-reducing enzymes (Fig. S3); FNR homologs identified in green sulfur bacteria were not identified in Nif-A-encoding genomes. Nif-A-encoding genomes tended to code for Fix, PFOR, and Rnf homologs, while Nif-B-encoding genomes tended to code for [FeFe]-hydrogenase, [NiFe]-hydrogenase (primarily group 4e) (41), and PFOR (Fig. S3). Genomes that coded for Nif-C were found to code for homologs of four of the seven Fd- or Fld-reducing enzymes, with [NiFe]-hydrogenase (primarily group 4e) (41) and PFOR being the most abundant among these. Nif-D-encoding genomes were found to code for only group 4d [NiFe]-hydrogenase (41) and PFOR. Anf- and Vnf-encoding genomes coded for homologs of five of the seven reducing enzymes. Fix, [FeFe]-hydrogenases, and PFOR were enriched in Anf-encoding genomes, while group 4e [NiFe]-hydrogenases (41), [FeFe]-hydrogenases, and PFOR were enriched in Vnf-encoding genomes.

The shift in the distribution of enzyme homologs capable of reducing Fd or Fld in diazotrophic genomes generally corresponded to the diversification of Nif-A from Nif-B/Nif-C/Nif-D (Fig. 1 and 3). Chemotrophic anaerobic diazotrophs that coded for Nif-B/Nif-C/Nif-D also encoded PFOR, [FeFe]-hydrogenase, or [NiFe]-hydrogenase. PFOR couples the oxidation of pyruvate to the reduction of Fd, and its expression is regulated by the availability of N in a variety of diazotrophs (36, 73, 100, 101), whereas specific lineages of hydrogenase (both [FeFe] and [NiFe]) have been shown to couple reversible H2 oxidation to the reduction of Fd (Fig. 3) (41, 102, 103).

Like chemotrophic anaerobes, anaerobic anoxygenic phototrophic green (both sulfur and nonsulfur) bacteria and anaerobic anoxygenic purple (both sulfur and nonsulfur) bacteria tended to encode PFOR (92% and 60% of total genomes, respectively). Anaerobic anoxygenic purple bacteria also encoded Fix, Rnf, [NiFe]-hydrogenase, and [FeFe]-hydrogenase, while green sulfur bacteria also encoded a functionally similar but phylogenetically distinct FNR isoform that is found in *Cyanobacteria* (Fig. 3). Like anaerobic diazotrophs, PFOR was identified in the genomes of aerobic/facultatively anaerobic diazotrophs. However, unlike anaerobic diazotrophs, the genomes of aerobic/facultatively anaerobic diazotrophs also tended to encode Fix, Rnf, and FNR (Fig. 3). Fix and Rnf generate reduced Fd with a more-reduced potential than its
substrate NADH via electron bifurcation (27, 56) or by coupling the reaction with reverse ion translocation (either proton or sodium dependent) (55–57), respectively. Fix has been suggested to play a role in supplying reductant to Nif in aerobic diazotrophs (27, 104–106). Consistent with this suggestion, Fix and Rnf are encoded in the genomes of aerobic chemotrophs (50% of these genomes) and anoxygenic phototrophs (68% of these genomes), specifically purple sulfur and nonsulfur bacteria (Fig. 3). Furthermore, homologs of FNR were only identified in the genomes of oxygenic phototrophic and anoxygenic green sulfur phototrophic bacteria, although the FNR homologs identified in green sulfur phototrophic bacteria were phylogenetically distinct from those cyanobacterial FNR. In Cyanobacteria it is thought that FNR catalyzes the reduction of Fd with NADPH generated during carbohydrate oxidation, which is then used to fix N₂ (61); it is not yet known if Fd produced by FNR in green sulfur bacteria is involved in N₂ fixation. Together, these observations suggest that the enzymes that diazotrophs use to generate reduced Fd or Fld to drive N₂ reduction vary in organisms with different metabolisms, in particular those that are able to integrate O₂ or light into their energy metabolism compared to those that are not able to take advantage of these metabolic strategies.

**Predicted electron transfer pathways to nitrogenase.** The covariation among the distributions of specified nitrogenase lineage homologs, Fd/Fld homologs, and Fd-/Fld-reducing-enzyme homologs was examined to identify putative systems of electron delivery to Nif in different metabolic backgrounds (Fig. 4). The primary Fd-/Fld-reducing-enzyme homologs in genomes that code for Nif-A are Fix, PFOR, and Rnf; FNR...
homologs were enriched in the genomes of Nif-A-encoding diazotrophic Cyanobacteria (n = 20). The source of electrons to Nif-A is likely either pyruvate in the case of PFOR (107), NADH generated from central metabolism in the case of Fix and Rnf (55, 56, 58), or NADPH in the case of FNR (60, 108).

Genomes that code for Nif-B were associated with different Fd/Fld homologs than Nif-A-encoding genomes and were associated with homologs of three different Fd-/Fld-reducing enzymes: PFOR, [FeFe]-hydrogenase, and [NiFe]-hydrogenase (Fig. 4). FNR was detected in all Nif-B-encoding green sulfur bacterial anoxygenic phototrophs (n = 11). To this end, the source of electrons to reduce Nif-B is predicted to be H2 in the case of [FeFe]- and [NiFe]-hydrogenases (109, 110), pyruvate in the case of PFOR, or NADPH in the case of FNR.

Genomes that encoded Nif-C and Nif-D coded for a different suite of Fd/Fld homologs potentially involved in electron delivery to Nif than did Nif-A- and Nif-B-encoding genomes (Fig. 4). Very little variation was observed in the distribution of putative Fd-/Fld-reducing-enzyme homologs and Fd/Fld homologs in Nif-C-encoding genomes compared to their distribution in genomes that encoded Nif-D. However, unlike Nif-C-encoding genomes, Nif-D-encoding genomes, which are derived from methanogens, lack evidence for homologs of [FeFe]-hydrogenase. This is consistent

**FIG 4** Bubble plot depicting the dominant patterns in the distribution of putative electron carrier protein homologs (Fds/Flds), enzyme homologs that putatively reduce Fd/Fld, and specified Nif lineages, as determined by the Apriori algorithm (124). Each unique pattern is given as a bubble, and the color represents the confidence value, or statistical significance, in the codistribution of specified proteins (only confidence values of ≥0.6 are presented). The size of the bubble represents the support value (≥0.2), or the frequency with which two proteins are identified in the same genome. For simplicity, only the proteins that were present in >20% of the diazotrophic genomes for each specified nitrogenase lineage were considered in this analysis. PFOR, pyruvate-Fld oxidoreductase; Rnf, *Rhodobacter* nitrogen fixation protein; FeFe, iron-only hydrogenase; NiFe, nickel-iron hydrogenase; FixABCX, electron transfer flavoprotein that is involved in nitrogen fixation. Names of source organisms for Fds and Flds are abbreviated as follows: Mv, *Methanocaldococcus vulcanius*; Mm, *Methanococcus maripaludis*; Mb, *Methanosarcina barkeri*; Cs, *Caldicellulosiruptor saccharolyticus*; Cp, *Clostridium pasteurianum*; Av, *A. vinelandii*. Abbreviations for Fds and Flds are presented in Tables 1 and 2, respectively.
with the absence of genes encoding [FeFe]-hydrogenase in Archaea (102, 103). Thus, like Nif-B, the source of electrons to reduce Nif-C and Nif-D is likely H₂ or pyruvate.

Anf- and Vnf-encoding diazotrophic genomes also encode Nif, which in the case of Anf was typically Nif-A and for Vnf was typically Nif-B. As such, patterns in the distribution of Fds/Flds and Fd/Fld-reducing enzymes in Anf- and Vnf-encoding genomes are like Nif-A and Nif-B, respectively. The organisms that encode Anf are likely to rely on three putative Fd-/Fld-reducing enzymes: [FeFe]-hydrogenase, Fix, or PFOR (Fig. S4). As in Nif-A-encoding organisms, the source of electrons to Anf is predicted to be from pyruvate or from central metabolism in the form of NADH. The organisms that encode Vnf are likely to rely on three putative Fd-/Fld-reducing enzymes: PFOR, [FeFe]-hydrogenase, and [NiFe]-hydrogenase (Fig. S4). As such, the source of electrons used to reduce Vnf is likely from H₂ or pyruvate.

Intriguingly, the genomes of several putative diazotrophs (n = 25) lacked a homolog of PFOR, [NiFe]-hydrogenase, [FeFe]-hydrogenase, Rnf, Fix, or either isoform of FNR (Table S2). The inability to detect homologs of these seven enzymes was not due to these genomes being incomplete, since all of them were closed. Of the 25 genomes that lacked a homolog of these enzymes, 14 encoded Nif-A-like nitrogenases, while 11 encoded Nif-B-like nitrogenases. All 14 genomes that encoded Nif-A-like nitrogenases were classified as either aerobic or facultatively anaerobic, while all 11 genomes that encoded Nif-B-like nitrogenases were classified as strictly anaerobic. These data suggest the presence of at least one additional enzymatic mechanism for generating Fds or Flds that can be used to reduce N₂.

Possible drivers in the diversification of pathways for electron transfer to nitrogenase. Previous studies have shown that Nif diversified in large part in response to O₂, in particular the integration of O₂ into cellular metabolism (17,111). By extension, this indicates that O₂ would have been at least temporarily available in the local habitat of the ancestors of aerobic/facultatively anaerobic diazotrophs, which would increase the oxidation state of their local habitat. Previous studies have noted a strong correlation between the average oxidation state of carbon in archaeal and bacterial proteomes (inferred from metagenomic data) and the oxidation state of the local environment (112). Likewise, extracellular proteins from yeast have been shown to have a higher average oxidation state of carbon than cytoplasmic proteins, an observation that was attributed to a higher oxidation state on the exterior of the cell than in the interior (113). As has been suggested previously, a shift in the oxidation state of a system to a more-oxidizing potential favors the formation of products (e.g., proteins) that themselves are more oxidized (113). Since the oxidation state of the cytoplasm of a cell should be related to that of the external environment of a cell, and cells that are under selection should be under selection to minimize the cost of protein synthesis to the extent that those proteins remain functional, we hypothesized that the integration of O₂ into the metabolism of diazotrophs would be accompanied by an overall shift in the oxidation state of the proteome of those cells that may in turn influence the functionality or stability of Fd/Flds, the enzymes that function to reduce these cofactors, or the availability of substrates for these enzymes. Observations supporting this hypothesis would serve as evidence that O₂ had a fundamental role in the evolution of diazotrophs, beyond what has been recognized previously, which includes physiological mechanisms of spatial/temporal decoupling of nitrogenase with O₂ metabolism (18, 19), increased consumption of O₂ through respiratory activity (20), and recruitment and loss of genes involved in regulating nitrogenase expression and activity (17).

To test this hypothesis, we calculated the average oxidation state of carbon in inferred proteomes for the taxa that comprised organisms encoding each of the four Nif sublineages (Fig. 5). Welch’s t test showed that the average oxidation state of carbon in inferred proteomes was significantly ($P \leq 2.2 \times 10^{-16}$) more positive ($-0.024 \pm 0.018$) for Nif-A-encoding genomes than for Nif-B-encoding genomes ($-0.043 \pm 0.019$), the most closely related lineage (Fig. 1). Likewise, the average oxidation state of carbon in inferred proteomes was significantly ($P = 1.9 \times 10^{-8}$ and $1.8 \times 10^{-8}$, respectively) more positive for Nif-A-encoding genomes than for genomes encoding Nif-C.
These observations are consistent with the observation that most of the Nif-A-encoding genomes were from aerobic or facultatively anaerobic taxa, whereas the majority of Nif-B-, Nif-C-, and Nif-D-encoding genomes were from strict anaerobes and, to a lesser extent, facultative anaerobes (Fig. 2A). These observations may be due to Nif-A-encoding diazotrophs inhabiting, on average, a more oxidized environment than Nif-B-, Nif-C-, and Nif-D-encoding diazotrophs. If true, this finding would suggest that Nif-A-encoding cells also may harbor a more oxidized cytoplasm than strictly anaerobic diazotrophs. A caveat to this analysis is that it assumes that all proteins in a diazotrophic cell are expressed at the same level, which is unlikely to be the case. Further experiments are needed to examine whether the average oxidation state of a cellular proteome is sensitive to growth conditions (e.g., aerobic versus anaerobic) and whether such a response reflects acclimatization to minimize energetic costs associated with protein biosynthesis.

We hypothesized that differences in the oxidation state of the external environment of cells and their cytoplasm would affect the functionality of proteins involved in electron transfer to nitrogenase and might account for the differences in the Fds/Flds used by aerobic facultative anaerobic and strictly anaerobic diazotrophs. To test this hypothesis, we reconstructed the evolutionary history of several Fds/Flds that appear to be commonly used as electron donors to the various lineages of Nif. The Fds FdxA and FdxB, which are commonly identified in Nif-A-encoding genomes, branch among (are nested within) a variety of Fd lineages commonly identified in anaerobic diazotrophs (i.e., Nif-B-, Nif-C-, and Nif-D-encoding taxa). These include MvFd4, MvFd5, CsFd6, MvFd2, and CpFd4 (Fig. S5). Parsimony would suggest that FdxA and FdxB are recently evolved and that these Fds evolved from Fds that likely functioned in an anaerobe. Similarly, NifF and FldA were nested among Flds commonly identified in anaerobic diazotroph genomes, including CpFld1, CpFld3, and CpFld2 (Fig. S6). Among Flds, NifF and FldA formed a monophyletic lineage and were sister to CpFld2, suggesting that they diverged from an ancestor of these proteins and that this ancestral protein most likely functioned in an anaerobe.

A previous study attempted to identify structural features associated with Fds from *Anabaena variabilis* that lend stability in oxic versus anoxic environments (114). The authors showed that slight changes in residues near the active-site FeS cluster protected it from reactive oxygen species. This indicates that the Fds that are active in...
aerobes inhabiting oxic environments likely have different amino acid compositions than Fds found in anaerobes, a finding that is supported by our informatics and phylogenetics analyses. As such, the results presented here provide a suite of Fds/Flds that appear to have evolved to function optimally in aerobic versus anaerobic diazotrophs and, thus, provide a template for downstream studies aimed at further elucidating the structural features that enable Fd/Fld function in more-oxidized environments.

The more-positive oxidation state of carbon in the inferred proteomes of Nif-A-encoding diazotrophs, which are largely from aerobic or facultatively anaerobic taxa, may indicate that the cytoplasm or local environment of those cells is more oxidizing. In general, H$_2$ (which itself has a low oxidation-reduction potential) would be expected to be at low abundance in an oxidized environment (112, 113, 115, 116). It follows that the diversification of Nif into more-oxidizing environments may have been accompanied by a decrease in available H$_2$, which in turn may have represented a selective pressure to evolve mechanisms to generate reduced Fd/Fld other than those that are dependent on H$_2$. In response, Nif-A-encoding diazotrophs would have continued to use PFOR, if available, to meet these demands. In other aerobic or facultatively anaerobic and anoxygenic phototrophic Nif-A-encoding diazotrophs, the primary electron carrier involved in central metabolism likely switched to NADH (51, 52). Under such conditions, diazotrophic cells would have been under selective pressure to evolve mechanisms to drive the formation of reduced Fd from NADH, such as Fix and Rnf.

**Concluding remarks.** Diazotrophs have evolved elegant mechanisms to overcome O$_2$ toxicity to nitrogenase during the transition from anaerobic to aerobic metabolism, including spatial and temporal decoupling of N$_2$ fixation activity from O$_2$ respiration (18, 19), increased O$_2$ respiration to maintain an anoxic cytoplasm (20), and recruitment and loss of genes involved in regulating nitrogenase expression and activity (17). Our data show that the average oxidation state of carbon in the inferred proteomes of diazotrophs also likely increased during the transition from anaerobic to aerobic metabolism, which would have driven large-scale changes in the composition of proteins and enzymes that drive the energy metabolism of diazotrophic cells, as well as the availability of substrates for those enzymes. Indeed, our analysis of the inferred systems of electron delivery to Nif reveals that substantial changes took place during the diversification of diazotrophs. These changes include those at the level of the primary electron donors that provide reductant to Nif, with early-evolving chemotrophic anaerobic diazotrophs likely supporting N$_2$ reduction with electrons derived from oxidation of H$_2$ or pyruvate. Later-evolving aerobic/facultatively anaerobic diazotrophs likely support this activity with electrons primarily in the form of NADH/NADPH derived from central metabolism or cyclic electron pathways in anoxygenic phototrophs. Similarly, a shift in the enzymes used to generate reduced Fd was observed, with chemotrophic anaerobic taxa inferred to be primarily dependent on [NiFe]-hydrogenase and PFOR and aerobic, facultative anaerobic or anoxygenic phototrophic taxa largely dependent on Fix and, to a lesser extent, PFOR and Rnf. Finally, a nearly complete turnover in the putative Fds/Flds in aerobic/facultatively anaerobic versus anaerobic diazotrophic taxa was observed. Collectively, these data are consistent with previous reports suggesting that O$_2$ had a profound influence on the evolution of nitrogenase (7, 13, 16, 17, 111) and further suggest that these changes may have been a global adaptive response to an increased oxidation state of the local environment.

The ability of cells to harness light energy likely also impacted systems of electron delivery to Nif. Due to the O$_2$ sensitivity of Nif, oxygenic phototrophs likely utilize FNR to generate reduced Fd from NADPH produced during carbohydrate fermentation at night or in specialized cells. The photosystems of anoxygenic green sulfur bacteria that also encode FNR energize electrons and shuttle them through a cyclic electron transport chain that involves Fd. It is possible that this Fd can be used to reduce Nif directly. Alternatively, these cells may reduce Fd with PFOR, which utilizes pyruvate produced from the oxidation of glycogen. In contrast, anoxygenic phototrophic bacteria, in
particular purple sulfur and nonsulfur bacteria, generate reduced quinones during light-driven electron transport, with those electrons supplied by the oxidation of exogenous inorganic or organic electron donors. The potential of these electrons can be further reduced through reverse electron transfer such that it is low enough to reduce NADH. These phototrophs often then utilize Fix or Rnf to drive the reduction of Fd with NADH.

The collective insights described herein suggest that the genomic and metabolic backgrounds of diazotrophs are associated with wholesale changes in the source of electron donors, enzymes involved in coupled oxidation of electron donors to the reduction of Fds/Flds, and the Fds/Flds used to reduce Nif. These changes were associated with differences in the ability to integrate O₂ and light into the energy metabolism of the cells. While the informatics-based data that are presented here are predictions of potential systems of electron delivery to Nif, in the case of Nif-A-encoding taxa, they are largely consistent with available biochemical data. However, the validity of the predictions made for pathways of electron delivery to Nif-B, Nif-C, and Nif-D-encoding taxa is unknown, since biochemical data have yet to be compiled for model taxa encoding homologs of these enzymes. These include diazotrophs that represent the most ancestral forms of Nif in both chemotrophic and phototrophic genomic and metabolic backgrounds.

MATERIALS AND METHODS

Identification and compilation of nitrogenase homologs. Genomes that encode the alpha (D) subunit of nitrogenase (i.e., NifD, AnfD and VnfD) were compiled as previously described (17, 117). Briefly, NifD homologs that were extracted using BLASTp were first aligned along with paralogs using Clustal Omega (118). The Nif-/Anf-/VnfD paralogs identified included ChlN from Chlamydomonas reinhardtii (ACJ50143) and Synechocystis sp. strain PCC 6803 (WP_010874227), BchN from Acidiphilium rubrum (BAA76536) and Chloroflexus aurantiacus (WP_012258416), and NifD from Methanocaldococcus jannaschii (WP_010870941) and Methanosarcina mazei (AAM30211). The aligned sequences were subjected to maximum-likelihood phylogenetic reconstruction with RAxML (version 7.3.0) (119), specifying the LG substitution matrix, the PROTGAMMA option, and 1,000 bootstrap iterations. Only Anf-/Vnf-/NifD protein homologs that clustered with previously characterized Nif-/Anf-/VnfD lineages (1, 13, 16) were extracted for downstream analysis. Furthermore, only genomes that encode the minimum set of proteins hypothesized to be required for N₂ fixation via Nif (i.e., nifHDKENB) or Anf/Vnf (i.e., anfHDKD/vnfHDKD) were retained, as previously described (7). In addition, genomes that encode homologs of just nifHDKENB were also retained, given that physiological studies suggest that organisms with this minimum gene complement can assimilate N₂ (11, 78). A total of 359 genomes from putative diazotrophs were compiled. To retrieve corresponding homologs of the iron protein (i.e., NifH, AnfH, and VnfH) and the beta subunit (i.e., NifK, AnfK, and VnfK), we performed a BLASTp search with NifH (ACO76403) and NifK (ACO76405) from Azotobacter vinelandii as queries, specifying cutoffs of 30% percent amino acid sequence identities and 60% sequence coverage. Only subunits that colocalized (e.g., nifHDK in an apparent operon) were retained. Each of the three nitrogenase subunits (i.e., H, D, and K subunits of Anf, Vnf, and Nif) was aligned individually with Clustal Omega (118), using the default settings, and the resultant alignment blocks were concatenated. The concatenated HDK sequences were subjected to phylogenetic reconstruction as described above.

Identification of homologs of Fd, Fld, and putative Fd- or Fld-reducing enzymes in the genomes of diazotrophs. To identify Fds and Flds in the genomes of putative diazotrophs, we first identified all the Fds and Flds in representative organisms for each Nif sublineage. These included Azotobacter vinelandii, Klebsiella pneumoniae, Rhodopseudomonas palustris, and Rhodobacter capsulatus for Nif-A, Clostridium pasteurianum and Methanosarcina barkeri for Nif-B, Caldicellulosiruptor saccharolyticus and Methanocaldococcus vulcanus for Nif-C, and Methanococcus maripaludis for Nif-D. All protein sequences that were annotated as Fds or Flds in the genomes of the above-named taxa were compiled. These protein sequences were subjected to Conserved Domain Database (120) BLAST to identify the type of iron-sulfur clusters present in the Fd and to confirm that the extracted Fld contained a flavin binding motif(s). All the compiled and curated Fds and Flds identified in these representative genomes were separately aligned using Clustal Omega (118). The aligned Fd or Fld sequences were subjected to maximum-likelihood phylogenetic reconstruction with RAxML (version 7.3.0) (119), specifying the LG substitution matrix and the PROTGAMMA option to empirically cluster the sequences into unique groups. In total, 48 Fds (Table 1) and five unique Flds (Table 2) were identified among diazotrophic genomes. The amino acid sequences of all the Fds and Flds were then used as bait sequences to identify homologs in the genomes of putative diazotrophs using BLASTp.

Putative enzymes that have been implicated in providing reductant to nitrogenase in the form of reduced Fd and/or Fld include the following: (i) pyruvate-flavodoxin oxidoreductase (PFOR) (35–38), (ii) ferredoxin-NADP⁺ oxidoreductase (FNR) (35, 62), (iii) Rhodobacter nitrogen fixation protein (Rnf) (81), and (iv) electron transfer flavoprotein (Fix) (27, 53, 54). Homologs of these enzymes were compiled from genomes of putative diazotrophs using BLASTp. The amino acid sequences of PFOR from Klebsiella
pneumoniae (WP_064371580) and Rhodopseudomonas palustris (CAE30161), FNR from Anabaena azollae (WP_013193003) and Chlorobium tepidum strain TLS (AAM72739), Rnf from Azotobacter vinelandii (ACO81179 to ACO81185), and Fix from Rhodopseudomonas palustris (WP_011665892 to WP_011665895) were used as query sequences. In addition, [FeFe]-hydrogenase has been suggested to provide reductant for N₂ fixation in Clostridium pasteurianum (39). Thus, the catalytic subunit of [FeFe] hydrogenase (HydA) from Chlamydomonas reinhardtii (AAL23572) was used as the query in a BLASTp search. Finally, group 3c, 3d, 4d, and 4e [NiFe]-hydrogenases catalyze the reversible reduction of Fd with H₂ (41, 42, 102). A group 3c representative ([NiFe]-hydrogenase from Methanothermobacter marburgensis [WP_013296467]) and a group 4 representative ([NiFe]-hydrogenase from Pyrococcus abyssi [WP_010867842.1]) were used as BLASTp queries to identify all group 3 and 4 [NiFe]-hydrogenase large-subunit sequences. Compiled sequences were aligned as described above and manually curated to include only those enzymes that had proximal and distal Cys-Cys pairs, the ligands for the [NiFe] active site (121, 122). Extracted sequences

### TABLE 1 Ferredoxin homologs identified in the genomes of putative diazotrophs

<table>
<thead>
<tr>
<th>Annotation</th>
<th>Protein accession no.</th>
<th>Type of cluster</th>
<th>Reference(s) or source</th>
</tr>
</thead>
<tbody>
<tr>
<td>asl2914</td>
<td>WP_013190310</td>
<td>2[4Fe-4S]</td>
<td>126</td>
</tr>
<tr>
<td>FdxH</td>
<td>WP_013190616</td>
<td>[2Fe-2S]</td>
<td>127</td>
</tr>
<tr>
<td>PetF</td>
<td>WP_013189829</td>
<td>[2Fe-2S]</td>
<td>128</td>
</tr>
<tr>
<td>FdxA</td>
<td>ACO80005</td>
<td>[3Fe-4S] [4Fe-4S]</td>
<td>129</td>
</tr>
<tr>
<td>Fdxl</td>
<td>ACO76607</td>
<td>[4Fe-4S]</td>
<td>130</td>
</tr>
<tr>
<td>FdxN</td>
<td>ACO81189</td>
<td>[2Fe-4S]</td>
<td>82, 131</td>
</tr>
<tr>
<td>VnfE</td>
<td>ACO76526</td>
<td>[4Fe-4S]</td>
<td>132</td>
</tr>
<tr>
<td>XylT</td>
<td>ACO77112</td>
<td>CXXXXCXXXXXCCXCX</td>
<td>130</td>
</tr>
<tr>
<td>AvfD1</td>
<td>AGK16019</td>
<td>[2Fe-2S]</td>
<td>This study</td>
</tr>
<tr>
<td>AvfD2</td>
<td>AGK16622</td>
<td>[2Fe-2S]</td>
<td>This study</td>
</tr>
<tr>
<td>AvfD3</td>
<td>AGK13276</td>
<td>2[4Fe-4S]</td>
<td>This study</td>
</tr>
<tr>
<td>CsFd1</td>
<td>ABP66178</td>
<td>2[4Fe-4S]</td>
<td>This study</td>
</tr>
<tr>
<td>CsFd2</td>
<td>ABP66040</td>
<td>2[4Fe-4S]</td>
<td>This study</td>
</tr>
<tr>
<td>CsFd3</td>
<td>ABP66582</td>
<td>2[4Fe-4S]</td>
<td>This study</td>
</tr>
<tr>
<td>CsFd4</td>
<td>ABP67884</td>
<td>2[4Fe-4S]</td>
<td>This study</td>
</tr>
<tr>
<td>CsFd5</td>
<td>ABP67132</td>
<td>2[4Fe-4S]</td>
<td>This study</td>
</tr>
<tr>
<td>CsFd6</td>
<td>ABP67141</td>
<td>2[4Fe-4S]</td>
<td>This study</td>
</tr>
<tr>
<td>CpD1</td>
<td>AJA47502</td>
<td>[2Fe-2S]</td>
<td>This study</td>
</tr>
<tr>
<td>CpD2</td>
<td>AJA47129</td>
<td>[2Fe-4S]</td>
<td>This study</td>
</tr>
<tr>
<td>CpD3</td>
<td>AJA49513</td>
<td>[2Fe-4S]</td>
<td>This study</td>
</tr>
<tr>
<td>CpD4</td>
<td>AJA46278</td>
<td>[4Fe-4S]</td>
<td>This study</td>
</tr>
<tr>
<td>CpD5</td>
<td>AJA49585</td>
<td>[2Fe-4S]</td>
<td>This study</td>
</tr>
<tr>
<td>KpF1</td>
<td>BAS34286</td>
<td>[2Fe-2S]</td>
<td>This study</td>
</tr>
<tr>
<td>KpF2</td>
<td>BAS37351</td>
<td>[2Fe-2S]</td>
<td>This study</td>
</tr>
<tr>
<td>KpF3</td>
<td>BAS33241</td>
<td>[2Fe-2S]</td>
<td>This study</td>
</tr>
<tr>
<td>KpF4</td>
<td>BAS35880</td>
<td>CXXCXXCC</td>
<td>This study</td>
</tr>
<tr>
<td>MvFd1</td>
<td>ACX73019</td>
<td>2[4Fe-4S]</td>
<td>This study</td>
</tr>
<tr>
<td>MvFd2</td>
<td>ACX72493</td>
<td>2[4Fe-4S]</td>
<td>This study</td>
</tr>
<tr>
<td>MvFd3</td>
<td>ACX73480</td>
<td>2[4Fe-4S]</td>
<td>This study</td>
</tr>
<tr>
<td>MvFd4</td>
<td>ACX72328</td>
<td>2[4Fe-4S]</td>
<td>This study</td>
</tr>
<tr>
<td>MvFd5</td>
<td>ACX73314</td>
<td>2[4Fe-4S]</td>
<td>This study</td>
</tr>
<tr>
<td>MvFd6</td>
<td>ACX73502</td>
<td>2[4Fe-4S]</td>
<td>This study</td>
</tr>
<tr>
<td>MvFd7</td>
<td>ACX72560</td>
<td>2[4Fe-4S]</td>
<td>This study</td>
</tr>
<tr>
<td>MvFd8</td>
<td>ACX72349</td>
<td>2[4Fe-4S]</td>
<td>This study</td>
</tr>
<tr>
<td>MnFd1</td>
<td>CAF29654</td>
<td>2[4Fe-4S]</td>
<td>This study</td>
</tr>
<tr>
<td>MnFd2</td>
<td>AKB59153</td>
<td>2[4Fe-4S]</td>
<td>This study</td>
</tr>
<tr>
<td>MnFd3</td>
<td>AKB57181</td>
<td>2[4Fe-4S]</td>
<td>This study</td>
</tr>
<tr>
<td>MnFd4</td>
<td>AKB57361</td>
<td>2[4Fe-4S]</td>
<td>This study</td>
</tr>
<tr>
<td>MnFd5</td>
<td>AKB58903</td>
<td>2[4Fe-4S]</td>
<td>This study</td>
</tr>
<tr>
<td>MnFd6</td>
<td>AKB59009</td>
<td>2[4Fe-4S]</td>
<td>This study</td>
</tr>
<tr>
<td>FdxC</td>
<td>ADE87009</td>
<td>[2Fe-25]</td>
<td>97, 133</td>
</tr>
<tr>
<td>FdxD</td>
<td>ADE84338</td>
<td>[2Fe-25]</td>
<td>80</td>
</tr>
<tr>
<td>FdxE</td>
<td>ADE86134</td>
<td>[2Fe-25]</td>
<td>93</td>
</tr>
<tr>
<td>FdxF</td>
<td>ABJ08445</td>
<td>[2Fe-4S]</td>
<td>81</td>
</tr>
<tr>
<td>Fer1</td>
<td>ADU46354</td>
<td>[4Fe-4S]</td>
<td>134</td>
</tr>
<tr>
<td>FerN</td>
<td>ACF03599</td>
<td>[4Fe-4S]</td>
<td>134</td>
</tr>
<tr>
<td>RpFd1</td>
<td>ABU04561</td>
<td>[2Fe-4S]</td>
<td>This study</td>
</tr>
</tbody>
</table>

aThe protein annotation, accession number for a representative protein homolog, inferred type of [FeS] cluster in a given protein homolog as predicted via the conserved domain database (120), and a literature reference, if available, are provided for each Fd identified. Av, A. vinelandii; Cs, Caldicellulosiruptor saccharolyticus; Cp, Clostridium pasteurianum; Kp, Klebsiella pneumoniae; Mv, Methanocaldococcus vulcanius; Mm, Methanococcus maripaludis; Mb, Methanosarcina barkeri; Rp, Rhodopseudomonas palustris.
The Apriori algorithm was used as implemented with the arules package in R (123, 124). For simplicity, only Fld-reducing enzymes in diazotrophic genomes. To identify the frequent patterns in this data set, the binary table was created based on the presence or absence of identified Fds, Flds, and Fd- or Fld-reducing enzymes.

### Table 2: Flavodoxin homologs identified in the genomes of putative diazotrophs

<table>
<thead>
<tr>
<th>Annotation</th>
<th>Accession no.</th>
<th>Reference or source</th>
</tr>
</thead>
<tbody>
<tr>
<td>NifF</td>
<td>ACO76434</td>
<td>36</td>
</tr>
<tr>
<td>CpFld1</td>
<td>AJA46461</td>
<td>This study</td>
</tr>
<tr>
<td>CpFld2</td>
<td>AJA47463</td>
<td>This study</td>
</tr>
<tr>
<td>CpFld3</td>
<td>AJA47660</td>
<td>This study</td>
</tr>
<tr>
<td>FldA</td>
<td>WP_011157672</td>
<td>93</td>
</tr>
</tbody>
</table>

The protein annotation, accession number for a representative protein homolog, and a literature reference, if available, is provided for each Fld identified. Cp, *Clostridium pasteurianum*.

Homologous to [NiFe]-hydrogenase homologs belonging to groups 3c, 3d, 4d, and 4e that have been proposed to be capable of reducing Fd were considered in this study (40–44).

**Identifying cooccurrence patterns in the distribution of Fds/Flds and putative Fd/Fld-reducing enzymes.** A binary table was created based on the presence or absence of identified Fds, Flds, and Fd- or Fld-reducing enzymes in diazotrophic genomes. To identify the frequent patterns in this data set, the Apriori algorithm was used as implemented with the arules package in R (123, 124). For simplicity, only the Fds, Flds, and Fd- or Fld-reducing enzymes present in >20% of the genomes were considered for this analysis. We applied the Apriori algorithm to our binary database while specifying the following parameters: a confidence threshold of 0.60 and a support threshold of 0.2. The confidence threshold is the measure of the statistical significance of an identified pattern within a data set, while the support threshold is indicative of the relative abundance of a given pattern in a data set. The Apriori algorithm works in two phases, with the first phase consisting of a scan of the entire database to objectively identify and extract frequently observed proteins. In the second phase, the algorithm uses those frequently occurring proteins to identify the dominant patterns in the distribution of proteins or combinations of those proteins. In doing so, the algorithm identifies correlations (within the user-defined confidence and support thresholds) between and among proteins in the database.

**Oxidation state of carbon.** The oxidation state of carbon for each amino acid in each protein sequence encoded in 359 putative diazotrophic genomes was calculated following the algorithm mentioned in reference 112 using a custom python script (version 2.7.12). The calculated oxidation number for each amino acid was then normalized to the total number of amino acids encoded in a genome to determine the average oxidation state of the proteome. This number was determined for each genome that encoded a specific Nif isomerase. These numbers were summed and then divided by the total number of genomes to arrive at the average oxidation state of carbon in proteomes inferred for genomes that encoded specific isofoms of nitrogenases (125).

**Supplemental Material**

Supplemental material for this article may be found at https://doi.org/10.1128/JB.00757-17.

**Supplemental File 1,** PDF file, 0.5 MB.

**Supplemental File 2,** XLSX file, 0.1 MB.

**Supplemental File 3,** XLSX file, 0.1 MB.

**Acknowledgments**

This work was supported as part of the Biological Electron Transfer and Catalysis Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, and Basic Energy Sciences under award number DE-SC0012518.

**References**


54. Ferredoxin and other iron-sulfur proteins. The formation of protein-heterocysts of the N2-fixing alga


