1 **Electron transfer to nitrogenase in different genomic and metabolic**

2 **backgrounds**

3

4 Saroj Poudel[1], Daniel R. Colman[1], Kathryn R. Fixen[2], Rhesa N. Ledbetter[4], Yanning Zheng[3],

5 Natasha Pence[5], Lance C. Seefeldt[4], John W. Peters[5], Caroline S. Harwood[3], and Eric S. Boyd[1,#]

6

7 [1]*Department of Microbiology and Immunology, Montana State University, Bozeman, Montana*

8 [2]*Department of Plant and Microbial Biology, University of Minnesota, St. Paul, Minnesota*

9 [3]*Department of Microbiology, University of Washington, Seattle, Washington*

10 [4]*Department of Chemistry and Biochemistry, Utah State University, Logan, Utah*

11 [5]*Institute of Biological Chemistry, Washington State University, Pullman, Washington*

12

13 Running Head (limit 54 characters and spaces): Electron delivery to nitrogenase

14

15 [#]Author of correspondence: Eric S. Boyd (*eboyd@montana.edu*)
16   Department of Microbiology & Immunology
17   109 Lewis Hall
18   Montana State University
19   Bozeman, MT 59717
20   Phone: (406) 994-7046
21   Fax: (406) 994-4926

22 **Keywords.**

23 Nitrogenase, nitrogen fixation, ferredoxin, flavodoxin, Fix, Rnf, bifurcation

24

25

ABSTRACT

Nitrogenase catalyzes the reduction of dinitrogen (N_2) using low potential electrons from ferredoxin (Fd) or flavodoxin (Fld) through an ATP dependent process. Since its emergence in an anaerobic chemoautotroph, this oxygen (O_2) 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 N_2 fixers (diazotrophs). Six distinct lineages of nitrogenase were identified and their distributions largely corresponded to differences in the host cells' ability to integrate O_2 or light into energy metabolism. Predicted pathways of electron transfer to nitrogenase in aerobes, facultative anaerobes, and phototrophs varied from those in anaerobes at the level 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 N_2 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.

49   **IMPORTANCE**

50

51   Nitrogen fixation supplies fixed nitrogen to cells from a variety of genomic and

52   metabolic backgrounds including those of aerobes, facultative anaerobes, chemotrophs, and

53   phototrophs. Here, using informatics approaches applied to genomic data, we show that

54   pathways of electron transfer to nitrogenase in metabolically diverse diazotrophic taxa have

55   diversified primarily in response to host cells' acquired ability to integrate O_2 or light into their

56   energy metabolism. Acquisition of two key enzyme complexes enabled aerobic and facultatively

57   anaerobic phototrophic taxa to generate electrons of sufficiently low potential to reduce

58   nitrogenase: the bifurcation of electrons via the Fix complex or the coupling of Fd reduction to

59   reverse ion translocation via the *Rhodobacter* nitrogen fixation (Rnf) complex.

60

61   **INTRODUCTION**

62

63   Nitrogenase is an oxygen sensitive enzyme that catalyzes the reduction of dinitrogen (N_2)

64   to ammonia (NH_3), accounting for nearly two-thirds of fixed nitrogen (N) on earth today (1, 2)

65   and thereby modulating the global fixed N supply (3, 4). Nitrogenase is comprised of two

66   components: A homodimeric iron protein (H subunit) and a heterotetrameric dinitrogenase

67   reductase protein complex (D and K subunits) (5, 6). Homologs of nitrogenase are widely

68   distributed among Bacteria but have only been identified in one group of taxa within the archaeal

69   Euryarchaeota phylum, methanogens. Nitrogenase homologs have not been identified among

70   Eukarya (7-13). Three types of nitrogenase have been described that are evolutionarily and

71   structurally related but differ in the metallic composition of their active site: Molybdenum (Mo)

72   containing nitrogenase (Nif), vanadium (V) containing nitrogenases (Vnf), and iron (Fe)

73   containing nitrogenases (Anf) (14-17).

74

75   Phylogenetic evidence indicates that Nif emerged prior to Anf and Vnf and that the

76   earliest evolving Nif lineages are from anaerobic, hyperthermophilic methanogens, implicating a

77   chemotrophic origin for N_2 fixation in an anoxic environment (7, 10, 14, 17, 18). N_2 fixing

78   microorganisms (i.e., diazotrophs) have since diversified to function in a wide variety of

79   genomic and metabolic backgrounds, including those of aerobes, facultative anaerobes, and

80   phototrophs. Given the O_2 sensitivity of Nif, aerobic, facultatively anaerobic, and oxygenic

81   phototrophic diazotrophs have had to evolve or acquire one of several mechanisms to mitigate

82   the toxicity of O_2 to this enzyme during N_2 fixation (18). These mechanisms include fixing N_2

83   only during dark hours when oxygenic photosynthesis ceases and when heterotrophic respiration

84   keeps O_2 tensions low (19), using specialized cells called heterocysts in filamentous

85   cyanobacteria to spatially localize nitrogenase in anaerobic cellular compartments (20),

86   maintaining an anoxic environment in the cytoplasm through increased O_2 dependent respiratory

87   activity (21), or fixing N_2 only during anaerobic growth. A recent study showed that

88   diversification of Nif in aerobes and facultative anaerobes was also accompanied by the

89   recruitment and loss of genes that are involved in regulating and protecting Nif against oxidative

90   stress (18).

91

92   The primary electron donors to Nif are reduced ferredoxin (Fd) (22-25) and flavodoxin

93   (Fld) (26-28). Transfer of eight electrons from Fd or Fld to NifH and ultimately to NifDK is an

94   ATP dependent process, requiring at a minimum 16 mol ATP per mol N_2 reduced (29-31). Less

95   is known of the stoichiometry of ATP hydrolysis per mol N_2 reduced by Anf or Vnf (32-35).

96   Reduction of Fd or Fld in diazotrophs that occupy anoxic environments (e.g., Clostridia and

97   methanogens) occurs via the activity of pyruvate-flavodoxin oxidoreductase (PFOR; (36-39)), a

98   select subset (group 3c, 3d, 4d, and 4e) of [NiFe]-hydrogenases, or [FeFe]-hydrogenase (40-46).

99   However, reduction of Fd (E_o' ~ -420 mV) in aerobic and some anoxygenic phototrophic

100   diazotrophs that inhabit less reducing environments is more of a challenge. Rather than

101   producing reduced Fd during their primary energy metabolism, aerobic bacteria and some

102   anoxygenic phototrophs, in particular facultatively anaerobic purple sulfur and non-sulfur

103   bacteria, generate reduced NADH or NADPH (E_o' = -320 mV) (47, 48), which are not of low

104   enough potential to drive N_2 reduction (49-51). Anaerobic purple sulfur and facultatively

105   anaerobic non-sulfur anoxygenic phototrophic bacteria utilize photosystems to drive cyclic

106   electron transfer. Reduction of $NAD^+/NADP^+$ is typically accomplished with electrons supplied

107   by oxidation of an inorganic substrate (e.g., sulfide, thiosulfate, H_2) or by the oxidation of

108   organic compounds and energy from reverse electron transport if the inorganic electron donor is

109   not of low enough potential to reduce $NAD^+/NADP^+$ (52, 53). It has been proposed that these

110   taxa acquired the Fix complex (encoded by *fixABCX*) and/or the *Rhodobacter* nitrogen fixation

111   [Rnf encoded by *rnfABCDEG*(*H*)] complex, in order to generate reduced Fd from

112   NADH/NADPH (47). Fix catalyzes the oxidation of two NADH to generate a reduced quinone,

113   which fuels the respiratory chain, and a reduced Fd (28, 54, 55) whereas Rnf catalyzes the

114   NADH-dependent reduction of Fd by coupling it to the depletion of the electrochemical gradient

115   (56-59).

116

117   Oxygenic phototrophic cyanobacteria utilize photosystem I to energize electrons to

118   potentials negative enough to drive reduction of Fd (52, 53). However, this Fd is not available to

119   Nif since it must be temporally or spatially separated from oxygenic photosynthesis due to

120   inhibition of Nif by O_2. Rather than code for [FeFe]-hydrogenase, [NiFe]-hydrogenase, Fix, or

121   Rnf and use these enzymes to reduce Fd, cyanobacteria encode ferredoxin-NADP$^+$

122   oxidoreductase (FNR) that can function in reverse to reduce Fd or Fld with NADPH generated

123   by carbohydrate oxidation in heterocysts or when O_2 tensions are low (36, 60-63). Some

124   cyanobacteria also encode for PFOR, which might be expected to contribute to Fd reduction in

125   these cells and this might be used by Nif. Like cyanobacteria, anaerobic anoxygenic green sulfur

126   bacteria use a Type I photosystem that is distantly related to photosystem I to generate reduced

127   Fd as a component of photosystem driven cyclic electron transfer (52, 53); however, unlike

128   cyanobacteria, it is possible that this may serve as reductant for N_2 fixation (64). Green sulfur

129   bacteria also encode an FNR that is phylogenetically and structurally unrelated to conventional

130   FNR found in cyanobacteria and this can also be used to drive reduction of Fd (65, 66). Thus, at

131   least seven enzyme complexes have evolved to provide reduced Fd for N_2 fixation: PFOR,

132   [NiFe]-hydrogenase, [FeFe]-hydrogenase, Rnf, Fix, and both forms of FNR (36); however, their

133   distribution in the genomes of diazotrophs is not known.

134

135   Different Fds and Flds are also likely to be involved in delivery of electrons to

136   nitrogenase in aerobes, anaerobes, and phototrophs, and may vary alongside the primary enzyme

137   that is involved in reducing these electron carriers. Fds are sensitive to O_2 due to the lability of

138   their iron sulfur (FeS) clusters (67-70). In contrast, Flds contain flavin mono-nucleotide as the

139   prosthetic group involved in electron transfers instead of FeS clusters and hence are thought to

140   be less sensitive to O_2 (68, 71, 72). Previous bioinformatics analyses have shown that NifF (18),

141   a Fld that can donate electrons to Nif (25, 73, 74), was recruited to *nif* operons during the

142   diversification of Nif from anaerobic to aerobic taxa (18), which may point to the use of Flds as

143   an adaptive strategy to fix N_2 in oxic environments. Moreover, under iron-deficient conditions

144   that characterize most circumneutral oxic environments, diazotrophs tend to synthesize Flds

145   preferentially as primary electron donors to Nif (23-25, 27, 75). Studies have also shown that

146   electron delivery by Fd or Fld can be complemented by other Fds or Flds that are encoded in the

147   genomes of diazotrophs (25, 36, 75-77). These observations suggest that pathways that mediate

148   electron flow to nitrogenase are flexible and vary according to the genomic and metabolic

149   background of taxa.

150

151   To better define the electron transfer system to nitrogenase in diverse microbes, we

152   compiled all Fd and Fld homologs in Nif-encoding genomes and classified them using homology

153   based methods. In addition, we compiled homologs of all enzymes that have been shown to

154   reduce Fd or Fld for use in the reduction of N_2 by nitrogenase. These include PFOR, [NiFe]-

155   hydrogenase, [FeFe]-hydrogenase, Fix, Rnf, and both forms of FNR. Since Nif is encoded in all

156   genomes that encode for Anf and Vnf (9), we also examined electron transfer systems to these

157   alternative nitrogenase isoforms. Organisms encoding for Nif, Anf, and Vnf were classified

158   phylogenetically and physiologically as aerobes, anaerobes, or facultative anaerobes and as

159   chemotrophs, anoxygenic phototrophs, or oxygenic phototrophs based on published data.

160   Statistical analyses were then applied to this curated dataset to identify patterns of co-occurrence

161   between the distribution of nitrogenase lineages/isoforms, enzymes that putatively reduce Fd/Fld,

162   and Fds/Flds. The results are discussed in the context of the metabolism of the cells, specifically

163   the influence of O_2 and light on putative electron delivery pathways to nitrogenase.

164

165   **RESULTS AND DISCUSSION**

166

167   ***Taxonomic distribution and phylogeny of HDK homologs.*** Nitrogenases were

168   identified in the genomes of diverse Bacteria and Archaea that included obligate aerobes,

169   facultative anaerobes, obligate anaerobes, phototrophs and chemotrophs, and autotrophs and

170   heterotrophs. The identification of Nif in organisms with diverse metabolisms is consistent with

171   a complex evolutionary history that has been described previously for N_2 fixation (7-9, 14, 17)

172   (**Supp. Table 1A**). Of the total 4588 publicly available genomes in our database, 359 genomes

173   (7.8% of total) encoded the minimal set of proteins for nitrogen fixation (i.e., homologs of

174   NifHDKENB) (7). Forty-six of the 359 taxa with genomes that encode for nitrogenase

175   complements have been experimentally shown to grow with atmospheric N_2 as their sole N

176   source (**Supp. Table 1B**). Of these 359 nitrogenase encoding genomes, 48 belonged to obligately

177   chemotrophic and anaerobic Archaea, all of which were affiliated with methanogens within the

178   phylum Euryarchaeota. The remaining 311 diazotrophic genomes were identified in the bacterial

179   domain, with the majority identified as members of the Proteobacteria (n=191) and Firmicutes

180   (n=68) (**Supp. Table 1A**). Of the 311 diazotrophic bacterial genomes, 31% were from aerobes,

181   40% were from facultative anaerobes, and 28% were from obligate anaerobes. Further, 79% of

182   the diazotrophic bacterial genomes were from chemotrophs and 21% were from phototrophs

183   (both oxygenic and anoxygenic).

184

185

186   Phylogenetic reconstruction of a concatenation of HDK proteins revealed six distinct

187   lineages of nitrogenases (**Fig. 1**). These included the four lineages that have been identified in

188   previous studies (8, 9, 14, 15, 17, 78), which include two Nif sub-lineages (designated as Nif-A

189   and Nif-B) and two sub-lineages designated as Anf and Vnf. In addition, previous combined

190   informatics and structural analyses have suggested that two biochemically uncharacterized

191   lineages likely harbor a molybdenum co-factor (10) and that organisms that encode them have

192   experimentally been shown to fix N_2 (12, 79), and hence we have referred to these lineages as

193   Nif-C and Nif-D in this study.

194

195   A total of 224 genomes encoded Nif-A and these were from taxa that are primarily from

196   the Proteobacteria (n=169), Firmicutes (n=26), and Cyanobacteria (n=20) (**Supp. Table 1A**).

197   Thirty of these taxa have been experimentally shown to fix N_2 (**Supp. Table 1B**). However, the

198   cyanobacterium *Microcoleus chthonoplastes* (which is not included in our database since a

199   complete genome is not available) does encode for Nif-A and the minimal set of proteins that

200   allow for N_2 fixation (i.e., homologs of NifHDKENB) (7). Yet cultivation studies suggest that

201   this taxon cannot grow with N_2 as its sole N source (80), indicating that the presence of full *nif*

202   gene complement does not guarantee the ability to fix N_2. The majority of the taxa that encode

203   for Nif-A are aerobes (42%) or facultative anaerobes (38% of total taxa) (**Fig. 2A**). In addition,

204   23% of Nif-A encoding genomes were from phototrophs and 77% were from chemotrophs

205   (**Supp. Fig. 1**). Among the 23% of Nif-A homologs that are encoded in the genomes of

206   phototrophs, 9% are from cyanobacteria, 11% are from facultatively anaerobic anoxygenic

207   purple non-sulfur bacteria, and 2% are from anaerobic anoxygenic purple bacteria. However, it is

208   not clear how rigorously O_2 usage has been characterized in these anaerobic anoxygenic purple

209   bacteria, which includes the following strains: *Allochromatium vinosum*, *Thiocystis violascens*,

210   *Thioflavicoccus mobilis*, and *Halorhodospira halophila*.

211

212   The majority of the 106 genomes that encode Nif-B were from obligate anaerobes (93%

213   of total taxa) and chemotrophs (87% of the total taxa) and anaerobic anoxygenic green

214   phototrophic bacteria (11% of total). Nif-B encoding taxa were primarily affiliated with

215   Firmicutes (n=34), Euryarchaeota (n=29), and Proteobacteria (n=22). Of these 106 Nif-B

216   encoding genomes, 13 were from taxa (primarily from the genus *Clostridium*) that have been

217   experimentally shown to fix N_2 (**Supp. Table 1B**). All of the anaerobic anoxygenic phototrophs

218   were affiliated with the Chlorobi or Chloroflexi (n=12). Members of the Firmicutes (n=8) and

219   Euryarchaeota (n=6) encoded Nif-C (n=15) while all (n=13) Nif-D encoding genomes were

220   methanogens (Euryarchaeota). None of the taxa that encode for Nif-C have been shown to fix N_2

221   fixation, while three organisms that encode for Nif-D (all of which are methanogens) have been

222   experimentally shown to fix N_2 (**Supp. Table 1B**). Taxa that encode Nif-C were primarily

223   anaerobes (88% of the total taxa) while taxa that encode Nif-D were all anaerobes. All the Nif-C

224   and Nif-D encoding genomes were chemotrophs.

225

226   A total of 32 genomes encoded Anf and these were primarily from Proteobacteria (n=16)

227   and Firmicutes (n=11) while a total of 23 genomes encoded Vnf. Most of the Vnf encoding

228   genomes were from the Euryarchaeota (n=11) and Firmicutes (n=8). A separate lineage

229   comprising HDK proteins from taxa that have not yet been shown to fix N_2 were nested among

230   Nif sub-lineages. These proteins were termed "uncharacterized" (abbreviated as Unc) as

10

231   previously described (14) and were affiliated with members of the Chloroflexi (i.e., *Roseiflexus*

232   spp.).

233

234   Multiple forms of nitrogenase were often detected in the same genome (**Fig. 2B**).

235   However, the genomes that code for Nif-C or Nif-D did not code for any other forms of

236   nitrogenases. The genome of the cyanobacterium *Pleurocapsa* sp. PCC 7327 was found to

237   encode both Nif-A and Nif-B. All Anf and Vnf encoding genomes also encoded Nif which is

238   consistent with previous observations (7, 9, 14). Four genomes were identified that coded for

239   Anf, Vnf, and Nif-B while two genomes encoded Anf, Vnf, and Nif-A. Interestingly, Anf was

240   more commonly detected in Nif-A encoding genomes (65% of total Anf encoding genomes)

241   while Vnf was more commonly detected in Nif-B encoding genomes (77% of total Vnf encoding

242   genomes). If the root of the nitrogenase phylogeny is within the Nif-D lineage, as has been

243   suggested previously (7, 14, 17), our phylogenetic reconstruction indicates that Anf diverged

244   from a Vnf-like ancestor, both of which emerged from a Nif-C or Nif-D-like ancestor. Likewise,

245   our phylogenetic reconstruction indicates that Nif-A diversified from a Nif-B-like ancestor.

246

247   ***Taxonomic distribution of ferredoxin and flavodoxin homologs in the genomes***

248   ***of putative diazotrophs.*** <u>*Ferredoxin (Fd).*</u> The taxonomic distribution of all 47 Fds identified in

249   the genomes of putative diazotrophs is detailed in **Supp. Table 2**. A total of 36 Fds were

250   detected in Nif-A encoding genomes, of which FdxB and FdxA were the most common and were

251   present in >40% of those genomes (**Supp. Fig. 2A**). *fdxB* is encoded near *nifQ* and *nifB* in the

252   genome of *A. vinelandii* (81, 82) and is expressed at a similar level to these genes under N_2-

253   fixing conditions (83). Mutational studies have shown involvement of FdxB in active site

11

254 metallocluster biosynthesis in *A. vinelandii* (84, 85). Further, FdxB was shown to be incapable of

255 serving as an electron donor to nitrogenase *in vitro* using protein from *Rhodobacter capsulatus*

256 (84). In contrast, in the cyanobacterium *Anabaena* sp. PCC 7120, FdxB has been shown to

257 complement *fdxH* mutant strains; FdxH is the preferred electron donor to nitrogenase in this

258 taxon (86). These observations suggest that FdxB has multiple roles in diazotrophic cells. The

259 prevalence of FdxA in Nif-A encoding genomes is consistent with its co-localization near major

260 *nif* gene clusters, such as in the case of *A. vinelandii* (25), or within the *nif* gene cluster itself,

261 such as in *Herbaspirillum seropedicae* (87). This agrees with findings from multiple studies that

262 have documented the ability of FdxA to donate electrons to nitrogenase (88-91). Several less

263 commonly detected Fds (identified in <40% of the Nif-A encoding genomes) have also been

264 shown or predicted to be involved in electron transfer to nitrogenase. These include FdxN in

265 *Rhodospirillum rubrum* and *Rhizobium meliloti* (92, 93), FdxE in *Rhodobacter capsulatus* (94),

266 FdxH (86, 95), FdxI in *Anabaena* sp. PCC 7120 (96, 97), and FdxC in *R. capsulatus* (98) (**Supp.**

267 **Fig. 2A**). Together, these results suggest that FdxA and FdxB may have a role in electron

268 delivery to Nif-A-like nitrogenases; however, other Fd could potentially complement the

269 functionality of these Fds.

270

271   The Fds CpFd1 (65% of the total Nif-B encoding genomes) and CpFd4 (48% of the total

272 genomes) correlated with the distribution of Nif-B in diazotrophic genomes (**Supp. Fig. 2A**),

273 suggesting a role for these Fds in electron delivery to Nif-B like nitrogenases. FdxA (47% of the

274 total Nif-C encoding genomes), CsFd3 (47% of the total genomes), CsFd6 (47% of the total

275 genomes), and CsFd1 (41% of the total genomes) were predominant in Nif-C encoding genomes

276 while Nif-D encoding genomes encoded multiple MvFds (**Supp. Fig. 2A**). The distribution of

12

277   Anf in genomes was only moderately correlated with the distribution of Fds, with FdxA and

278   FdxB yielding the highest correlations (i.e. >45% of the genome) (**Supp. Fig. 2A**). Lastly, the

279   distribution of Vnf was highly correlated with the distribution of CpFd4 (80% of the total

280   genomes) and to a lesser extent, MbFd1, MbFd2, MbFd3, and MbFd4 (**Supp. Fig. 2A**). While

281   electron transfer to Nif-A like nitrogenase has been extensively studied, we are unaware of

282   experimental data on the role of Fds in electron transfer to Nif-B, Nif-C, Nif-D, Anf, or Vnf,

283   which precludes a comparison of our informatics-based predictions to experimental data.

284

285   The Fds identified in our study varied markedly among genomes coding for the various

286   isoforms of nitrogenase (**Supp. Fig. 2A**). Importantly, genomes encoding Nif-A nitrogenases

287   coded for unique Fds that were either absent or present in rarely detected in the genomes of other

288   diazotrophs (**Supp. Fig. 2A**). For example, of 14 abundant Fds (i.e., presence in >40% of the

289   genomes) encoded by Nif-A encoding genomes, 12 were unique to Nif-A encoding genomes

290   while only two were identified in the genomes of other diazotrophs (**Supp. Fig. 2A**). In contrast,

291   considerable overlap was observed in the composition of Fds/Flds encoded in genomes that also

292   encoded for Nif-B, Nif-C, or Nif-D.

293

294   _Flavodoxin (Fld)_. We identified a total of five phylogenetically distinct Flds among

295   nitrogenase encoding genomes. The taxonomic distribution of all five identified Flds is detailed

296   in **Supp. Table 2.** Of the five Flds detected in Nif-A encoding genomes, FldA was the most

297   common and was identified among 26% of the genomes (**Supp. Fig. 2B**). The distribution of

298   CpFld2 and CpFld3 co-occurred with Nif-B in >40% of the genomes while CpFld3 was detected

299   in >40% of the Nif-C and Nif-D encoding genomes (**Supp. Fig. 2B**). Like Nif-A encoding

13

300   genomes, FldA was the dominant Fld in Anf encoding genomes (**Supp. Fig. 2B**). Lastly, CpFld3

301   and CpFld2 were the dominant Fld in Vnf encoding genomes (**Supp. Fig. 2B**). There is no

302   experimental evidence that these Flds are involved in electron transfer to nitrogenase with the

303   exception of NifF (present in a few Nif-A encoding genomes), which has been shown to transfer

304   electrons to nitrogenase in *Klebsiella pneumoniae* (99, 100) and *A. vinelandii* (25, 28).

305

306   The distribution of the five Flds identified in this study varied in diazotrophs with

307   different metabolic backgrounds. Of the five Flds identified among the genomes of diazotrophs,

308   NifF and FldA were only detected in the genomes of aerobes, facultative anaerobes, or

309   phototrophs that encoded Nif-A (**Supp. Fig. 2B**). The sole exception to this observation was in

310   the genome of the anaerobic spirochaete *Spirochaeta smaragdinae* DSM 11293, which was

311   found to code for FldA and Nif-B (**Supp. Table 2**). The remaining three Flds were common in

312   the genomes of anaerobic chemotrophs that encoded Nif-B, Nif-C, and Nif-D. This observation

313   is like that made for Fds, where the distribution of Flds in genomes that encode Nif-A were

314   distinct from those that encoded Nif-B, Nif-D, and Nif-D. These collective observations indicate

315   that the Fds/Flds in aerobic/facultatively anaerobic diazotrophs differ from those in anaerobic

316   diazotrophs, leading to the hypothesis that O_2 played a role in the diversification of electron

317   delivery systems to nitrogenase. Moreover, the Fds/Flds are different among phototrophs and

318   chemotrophs, leading to the hypothesis that integration of light into the energy metabolism of

319   diazotrophs played a key role in the diversification of electron delivery systems to nitrogenase

320

321   ***Distribution of ferredoxin and flavodoxin reducing protein homologs in the***

322   ***genomes of putative diazotrophs.*** The taxonomic distribution of PFOR, [NiFe]-hydrogenase,

323   [FeFe]-hydrogenase, Rnf, Fix, and FNR homologs are detailed in **Supp. Table 2**. Briefly,

324   genomes that encoded Nif-A and Nif-B coded for homologs of six out of the seven putative Fd

325   or Fld reducing enzymes (**Supp. Fig. 3**); FNR homologs identified in green sulfur bacteria were

326   not identified in Nif-A encoding genomes. Nif-A encoding genomes tended to code for Fix,

327   PFOR, and Rnf homologs while Nif-B encoding genomes tended to code for [FeFe]-

328   hydrogenase, [NiFe]-hydrogenase (primarily group 4e (42)), and PFOR (**Supp. Fig. 3**). Genomes

329   that coded for Nif-C were found to code for homologs of four of the seven Fd or Fld reducing

330   enzymes with [NiFe]-hydrogenase (primarily group 4e (42)) and PFOR being the most abundant

331   among these. Nif-D encoding genomes were found to code for only group 4d [NiFe]-

332   hydrogenase (42) and PFOR. Anf and Vnf encoding genomes coded for homologs of five of the

333   seven reducing enzymes. Fix, [FeFe]-hydrogenases, and PFOR were enriched in Anf encoding

334   genomes while group 4e [NiFe]-hydrogenases (42), [FeFe]-hydrogenases, and PFOR were

335   enriched in Vnf encoding genomes.

336

337   The shift in the distribution of enzyme homologs capable of reducing Fd or Fld in

338   diazotrophic genomes generally corresponded to the diversification of Nif-A from Nif-B/Nif-

339   C/Nif-D (**Figs. 1 & 3**). Chemotrophic anaerobic diazotrophs that coded for Nif-B/Nif-C/Nif-D

340   also encoded PFOR, [FeFe]-hydrogenase, or [NiFe]-hydrogenase. PFOR couples the oxidation

341   of pyruvate to the reduction of Fd and its expression is regulated by the availability of N in a

342   variety of diazotrophs (37, 74, 101, 102), whereas specific lineages of hydrogenase (both [FeFe]

343   and [NiFe]) have been shown to couple reversible H_2 oxidation to the reduction of Fd (**Fig. 3**)

344   (42, 103, 104).

345

346   Like chemotrophic anaerobes, anaerobic anoxygenic phototrophic green (both sulfur and

347   non-sulfur) bacteria and anaerobic anoxygenic purple (both sulfur and non-sulfur) bacteria

348   tended to encode PFOR (92% and 60% of total genomes, respectively). Anaerobic anoxygenic

349   purple bacteria, also encoded Fix, Rnf, [NiFe]-hydrogenase, and [FeFe]-hydrogenase while green

350   sulfur bacteria also encoded a functionally similar but phylogenetically distinct FNR isoform that

351   is found in cyanobacteria (**Fig. 3**). Like anaerobic diazotrophs, PFOR was identified in the

352   genomes of aerobic/facultatively anaerobic diazotrophs. However, unlike anaerobic diazotrophs,

353   the genomes of aerobic/facultatively anaerobic diazotrophs also tended to encode Fix, Rnf, and

354   FNR (**Fig. 3**). Fix and Rnf generate reduced Fd with a more reduced potential than its substrate

355   NADH via electron bifurcation (28, 57) or by coupling the reaction with reverse ion

356   translocation (either proton- or sodium-dependent) (56-58), respectively. Fix has been suggested

357   to play a role in supplying reductant to Nif in aerobic diazotrophs (28, 105-107). Consistent with

358   this suggestion, Fix and Rnf are encoded in the genomes of aerobic chemotrophs (50% of these

359   genomes) and anoxygenic phototrophs (68% of these genomes), specifically purple sulfur and

360   non-sulfur bacteria (**Fig. 3**). Further, homologs of FNR were only identified in the genomes of

361   oxygenic phototrophic and in anoxygenic green sulfur phototrophic bacteria, although FNR

362   homologs identified in green sulfur phototrophic bacteria were phylogenetically distinct from

363   those cyanobacterial FNR. In cyanobacteria it is thought that FNR catalyzes the reduction of Fd

364   with NADPH generated during carbohydrate oxidation which is then used to fix N_2 (62); it is not

365   yet known if Fd produced by FNR in green sulfur bacteria is involved in N_2 fixation. Together,

366   these observations suggest that the enzymes that diazotrophs use to generate reduced Fd or Fld to

367   drive N_2 reduction vary in organisms with different metabolisms, in particular those that are able

368   to integrate O_2 or light into their energy metabolism when compared to those that are not able to

369   take advantage of these metabolic strategies.

370

371   ***Predicted electron transfer pathways to nitrogenase.*** The co-variation among the

372   distribution of specified nitrogenase lineage homologs, Fd/Fld homologs, and Fd/Fld reducing

373   enzyme homologs was examined to identify putative electron delivery systems to Nif in different

374   metabolic backgrounds (**Fig. 4**). The primary Fd/Fld reducing enzyme homologs in genomes that

375   code for Nif-A are Fix, PFOR, and Rnf; FNR homologs were enriched in the genomes of Nif-A

376   encoding diazotrophic cyanobacteria (n=20). The source of electrons to Nif-A is likely either

377   pyruvate in the case of PFOR (108), NADH generated from central metabolism in the case of Fix

378   and Rnf (56, 57, 59), or NADPH in the case of FNR (61, 109).

379

380   Genomes that code for Nif-B were associated with different Fds/Flds homologs than Nif-

381   A encoding genomes and were associated with homologs of three different Fd/Fld reducing

382   enzymes: PFOR, [FeFe]-hydrogenase, and [NiFe]-hydrogenase (**Fig. 4**). FNR was detected in all

383   Nif-B encoding green sulfur bacterial anoxygenic phototrophs (n=11). To this end, the source of

384   electrons to reduce Nif-B is predicted to be H_2 in the case of [FeFe]- and [NiFe]-hydrogenase

385   (110, 111), pyruvate in the case of PFOR, or NADPH in the case of FNR.

386

387   Genomes that encoded Nif-C and Nif-D coded for a different suite of Fd/Fld homologs

388   potentially involved in electron delivery to Nif, when compared to Nif-A and Nif-B encoding

389   genomes (**Fig. 4**). Very little variation was observed in the distribution of putative Fd/Fld

390   reducing enzyme homologs and Fd/Fld homologs in Nif-C encoding genomes when compared to

391   genomes that encoded Nif-D. However, unlike Nif-C encoding genomes, Nif-D encoding

392   genomes, which are derived from methanogens, lack evidence for homologs of [FeFe]-

393   hydrogenase. This is consistent with the absence of genes encoding [FeFe]-hydrogenase in

394   Archaea (103, 104). Thus, like Nif-B, the source of electrons to reduce Nif-C and Nif-D is likely

395   H_2 or pyruvate.

396

397   Anf and Vnf encoding diazotrophic genomes also encode Nif, which in the case of Anf

398   was typically Nif-A, and for Vnf, typically Nif-B. As such, patterns in the distribution of Anf and

399   Vnf, Fds/Flds, and Fd/Fld reducing enzyme are like Nif-A and Nif-B, respectively. The

400   organisms that encode Anf are likely to rely on three putative Fd/Fld reducing enzymes: [FeFe]-

401   hydrogenase, Fix, or PFOR (**Supp. Fig. 4**). Like Nif-A encoding organisms, the source of

402   electrons to Anf is predicted to be from pyruvate or from central metabolism in the form of

403   NADH. The organisms that encode Vnf are likely to rely three putative Fd/Fld reducing

404   enzymes: PFOR, [FeFe]-hydrogenase, and [NiFe]-hydrogenase (**Supp. Fig. 4**). As such, the

405   source of electrons used to reduce Vnf is likely from H_2 or pyruvate.

406

407   Intriguingly, the genomes of several putative diazotrophs (n=25) lacked a homolog of

408   PFOR, [NiFe]-hydrogenase, [FeFe]-hydrogenase, Rnf, Fix, or both isoforms of FNR (**Supp.**

409   **Table 2**). The inability to detect homologs of these seven enzymes was not due to these genomes

410   being incomplete, since all of them were closed. Of the 25 genomes that lacked a homolog of

411   these enzymes, 14 encoded Nif-A-like nitrogenases while 11 encoded Nif-B-like nitrogenases.

412   All 14 genomes that encoded Nif-A-like nitrogenases were either classified as aerobic or

413   facultatively anaerobic while all 11 genomes that encoded Nif-B-like nitrogenases were

18

414 classified as strictly anaerobic. These data suggest the presence of at least one additional

415 enzymatic mechanism of generating Fds or Flds that can be used to reduce N_2.

416

417 ***Possible drivers in the diversification of electron transfer pathways to***

418 ***nitrogenase.*** Previous studies have shown that Nif diversified in large part in response to O_2, in

419 particular the integration of O_2 into cellular metabolism (18, 112). By extension, this indicates

420 that O_2 would have been at least temporarily available in the local habitat of the ancestors of

421 aerobic/facultatively anaerobic diazotrophs, which would increase the oxidation state of their

422 local habitat. Previous studies have noted a strong correlation between the average oxidation

423 state of carbon in archaeal and bacterial proteomes (inferred from metagenomic data) and the

424 oxidation state of the local environment (113). Likewise, extracellular proteins from yeast have

425 been shown to have a higher average oxidation state of carbon than cytoplasmic proteins, an

426 observation that was attributed to a higher oxidation state on the exterior of the cell when

427 compared to the interior (114). As has been suggested previously, a shift in the oxidation state of

428 a system to a more oxidizing potential favors formation of products (e.g., proteins) that

429 themselves are more oxidized (114). Since the oxidation state of the cytoplasm of a cell should

430 be related to that of the external environment of a cell, and cells that are under selection should

431 be under selection to minimize the cost of protein synthesis to the extent that those proteins

432 remain functional, we hypothesized that the integration of O_2 into the metabolism of diazotrophs

433 would be accompanied by an overall shift in the oxidation state of the proteome of those cells

434 which may in turn influence the functionality or stability of Fd/Flds, the enzymes that function to

435 reduce these co-factors, or the availability of substrates for these enzymes. Observations

436 supporting this hypothesis would serve as evidence that O_2 had a fundamental role in the

19

437 evolution of diazotrophs, beyond what has been recognized previously which includes

438 physiological mechanisms of spatial/temporal decoupling of nitrogenase with O_2 metabolism

439 (19, 20), increased consumption of O_2 through respiratory activity (21), and recruitment and loss

440 of genes involved in regulating nitrogenase expression and activity (18).

441

442   To test this hypothesis, we calculated the average oxidation state of carbon in inferred

443 proteomes for the taxa that comprised each of the four Nif sub-lineages (**Fig. 5**). A Welch's t-test

444 showed that the average oxidation state of carbon in inferred proteomes of Nif-A encoding

445 genomes was significantly (P <2.2 x 10^{-16}) more positive (-0.024 $\pm$ 0.018) than those for Nif-B (-

446 0.043 $\pm$ 0.019), the most closely related lineage (**Fig. 1**). Likewise, the average oxidation state of

447 carbon in inferred proteomes of Nif-A encoding genomes was significantly (P = 1.9 x 10^{-8} and

448 1.8 x 10^{-8}, respectively) more positive than that of Nif-C (-0.078 $\pm$ 0.033) and Nif-D (-0.058 $\pm$

449 0.016) encoding genomes. These observations are consistent with the observation that most of

450 the Nif-A encoding genomes were from aerobic or facultatively anaerobic taxa, whereas the

451 majority of Nif-B, Nif-C, and Nif-D encoding genomes were from strict anaerobes and to a

452 lesser extent, facultative anaerobes (**Fig. 2A**). These observations may be due to Nif-A encoding

453 diazotrophs inhabiting a more oxidized environment on average when compared to Nif-B, Nif-C,

454 and Nif-D encoding diazotrophs. If true, this finding would suggest that Nif-A encoding cells

455 also may harbor a more oxidized cytoplasm than strictly anaerobic diazotrophs. A caveat to this

456 analysis is that it assumes that all proteins in a diazotrophic cell (i.e., proteome) are expressed at

457 the same level, which is unlikely to be the case. Further experiments are needed to examine

458 whether the average oxidation state of a cellular proteome is sensitive to growth conditions (e.g.,

459   aerobic versus anaerobic) and whether such a response reflects acclimatization to minimize

460   energetic costs associated with protein biosynthesis.

461

462   We hypothesized that differences in the oxidation state of the external environment of

463   cells and their cytoplasm would affect the functionality of proteins involved in electron transfer

464   to nitrogenase and may account for the differences in the Fd/Flds used by aerobic/facultative

465   anaerobic and strictly anaerobic diazotrophs. To test this hypothesis, we reconstructed the

466   evolutionary history of several Fds/Flds that appear to be commonly used as electron donors to

467   the various lineages of Nif. The Fds FdxA and FdxB, which are commonly identified in Nif-A

468   encoding genomes, branch among (are nested within) a variety of Fd lineages commonly

469   identified in anaerobic diazotrophs (i.e., Nif-B, Nif-C, and Nif-D encoding taxa). These include

470   MvFd4, MvFd5, CsFd6, MvFd2, and CpFd4 (**Supp. Fig. 5**). Parsimony would suggest that FdxA

471   and FdxB are recently evolved and that these Fds evolved from Fds that likely functioned in an

472   anaerobe. Similarly, NifF and FldA were nested among Flds commonly identified in anaerobic

473   diazotroph genomes, including CpFld1, CpFld3, and CpFld2 (**Supp. Fig. 6**). Among Flds, NifF

474   and FldA formed a monophyletic lineage and were sister to CpFld2, suggesting that they

475   diverged from an ancestor of these proteins and that this ancestral protein most likely functioned

476   in an anaerobe.

477

478   A previous study attempted to identify structural features associated with Fds from

479   *Anabaena variabilis* that lend stability in oxic versus anoxic environments (115). The authors

480   showed that slight changes in residues near the active site FeS cluster protected it from reactive

481   oxygen species. This indicates that the Fds that are active in aerobes inhabiting oxic

482   environments likely have different amino-acid compositions than Fds found in anaerobes, a

483   finding that is supported by our informatics and phylogenetics analyses. As such, the results

484   presented here provide a suite of Fds/Flds that appear to have evolved to function optimally in

485   aerobic versus anaerobic diazotrophs and thus provide a template for downstream studies aimed

486   at further elucidating the structural features that enable Fd/Fld function in more oxidized

487   environments.

488

489   The more positive oxidation state of carbon in the inferred proteomes of Nif-A encoding

490   diazotrophs, which are largely from aerobic or facultatively anaerobic taxa, may indicate that the

491   cytoplasm or local environment of those cells is more oxidizing. In general, H_2 (which itself has

492   a low oxidation-reduction potential) would be expected to be of low abundance an oxidized

493   environment (113, 114, 116, 117). It follows that diversification of Nif into more oxidizing

494   environments may have been accompanied by a decrease in available H_2, which in turn may have

495   represented a selective pressure to evolve mechanisms to generate reduced Fd/Fld other than

496   those that are dependent on H_2. In response, Nif-A encoding diazotrophs would have continued

497   to use PFOR, if available, to meet these demands. In other aerobic or facultatively anaerobic and

498   anoxygenic phototrophic Nif-A encoding diazotrophs, the primary electron carrier involved in

499   central metabolism likely switched to NADH (52, 53). Under such conditions, diazotrophic cells

500   would have been under selective pressure to evolve mechanisms to drive formation of reduced

501   Fd from NADH, such as Fix and Rnf.

502

503   ***Concluding remarks.*** Diazotrophs have evolved elegant mechanisms to overcome O_2

504   toxicity to nitrogenase during the transition from anaerobic to aerobic metabolism including

22

505 spatial and temporal decoupling of N_2 fixation activity from O_2 respiration (19, 20), increased O_2

506 respiration to maintain an anoxic cytoplasm (21), and recruitment and loss of genes involved in

507 regulating nitrogenase expression and activity (18). Our data shows that the average oxidation

508 state of carbon in the inferred proteomes of diazotrophs also likely increased during the transition

509 from anaerobic to aerobic metabolism, which would have driven largescale changes in the

510 composition of proteins and enzymes that drive the energy metabolism of diazotrophic cells, as

511 well as the availability of substrates for those enzymes. Indeed, our analysis of the inferred

512 electron delivery systems to Nif reveals that substantial changes took place during the

513 diversification of diazotrophs. These changes include those at the level of the primary electron

514 donors that provide reductant to Nif, with early evolving chemotrophic anaerobic diazotrophs

515 likely supporting N_2 reduction with electrons derived from oxidation of H_2 or pyruvate. Later

516 evolving aerobic/facultatively anaerobic diazotrophs likely support this activity with electrons

517 primarily in the form of NADH/NADPH derived from central metabolism or cyclic electron

518 pathways in anoxygenic phototrophs. Similarly, a shift in the enzymes used to generate reduced

519 Fd was observed, with chemotrophic anaerobic taxa inferred to be primarily dependent on

520 [NiFe]-hydrogenase and PFOR and aerobic, facultative anaerobic or anoxygenic phototrophic

521 taxa largely dependent on Fix and to a lesser extent PFOR and Rnf. Finally, a near complete

522 turnover in the putative Fds/Flds in aerobic/facultatively anaerobic versus anaerobic diazotrophic

523 taxa was observed. Collectively, these data are consistent with previous reports suggesting that

524 O_2 had a profound influence on the evolution of nitrogenase (7, 14, 17, 18, 112) and further

525 suggest that these changes may have been a global adaptive response to an increased oxidation

526 state of the local environment.

527

528   The ability of cells to harness light energy likely also impacted electron delivery systems

529 to Nif. Due to O_2 sensitivity of Nif, oxygenic phototrophs likely utilize FNR to generate reduced

530 Fd from NADPH produced during carbohydrate fermentation at night or in specialized cells. The

531 photosystems of anoxygenic green sulfur bacteria that also encode FNR energize electrons and

532 shuttle them through a cyclic electron transport chain that involves Fd. It is possible that this Fd

533 can be used to reduce Nif directly. Alternatively, these cells may reduce Fd with PFOR that is

534 utilizing pyruvate produced from the oxidation of glycogen. In contrast, anoxygenic phototrophic

535 bacteria, in particular purple sulfur and non-sulfur bacteria, generate reduced quinones during

536 light driven electron transport, with those electrons supplied by oxidation of exogenous inorganic

537 or organic electron donors. The potential of these electrons can be further reduced through

538 reverse electron transfer such that they are low enough to reduce NADH. These phototrophs

539 often then utilize Fix or Rnf to drive the reduction of Fd with NADH.

540

541   The collective insights described herein suggest that the genomic and metabolic

542 background of diazotrophs are associated with wholesale changes in the source of electron

543 donors, enzymes involved in coupled oxidation of electron donors to reduction of Fds/Flds, and

544 the Fds/Flds used to reduce Nif. These changes were associated with differences in the ability to

545 integrate O_2 and light into the energy metabolism of the cells. While the informatics based data

546 that are presented here are predictions of potential electron delivery systems to Nif, in the case of

547 Nif-A encoding taxa, they are largely consistent with available biochemical data. However, the

548 validity of the predictions made for pathways of electron delivery to Nif-B, Nif-C, and Nif-D

549 encoding taxa are unknown, since biochemical data has yet to be compiled for model taxa

550 encoding homologs of these enzymes. These include diazotrophs that represent the most

551   ancestral forms of Nif in both chemotrophic and phototrophic genomic and metabolic

552   backgrounds.

553

554   **MATERIALS AND METHODS**

555

556   **Identification and compilation of nitrogenase homologs.** Genomes that encode the

557   alpha (D) subunit of nitrogenase (i.e., NifD, AnfD and VnfD) were compiled as previously

558   described (18, 118). Briefly, NifD homologs that were extracted using BLASTp were first

559   aligned along with paralogs using Clustal Omega (119). The Nif/Anf/VnfD paralogs identified

560   included ChlN from *Chlamydomonas reinhardtii* (ACJ50143) and *Synechocystis* sp. PCC 6803

561   (WP_010874227), BchN from *Acidiphilium rubrum* (BAA76536) and *Chloroflexus aurantiacus*

562   (WP_012258416), and NflD from *Methanocaldococcus jannaschii* (WP_010870941) and

563   *Methanosarcina mazei* (AAM30211). The aligned sequences were subjected to maximum-

564   likelihood phylogenetic reconstruction with RAxML (version 7.3.0) (120), specifying the LG

565   substitution matrix, the PROTGAMMA option and 1000 bootstraps iterations. Only

566   Anf/VnfD/NifD protein homologs that clustered with previously characterized Nif/Anf/VnfD

567   lineages (1, 121, 122) were extracted for downstream analysis. Furthermore, only genomes that

568   encode the minimum set of proteins hypothesized to be required for N_2 fixation via Nif (i.e.,

569   *nifHDKENB*) or Anf/Vnf (i.e., *anfHDK/vnfHDK*) were retained, as previously described (7). In

570   addition, genomes that encode for homologs of just *nifHDKEB* were also retained, given that

571   physiological studies suggest that organisms with this minimum gene complement can assimilate

572   N_2 (12, 79). A total of 359 genomes from putative diazotrophs were compiled. To retrieve

573   corresponding homologs of the iron protein (i.e., NifH, AnfH, and VnfH) and the beta subunit

25

574   (i.e., NifK, AnfK, and VnfK), we performed a BLASTp search with NifH (ACO76403) and

575   NifK (ACO76405) from *Azotobacter vinelandii* as queries, specifying cutoffs of 30% percent

576   amino acid sequence identities and 60% sequence coverage. Only subunits that co-localized

577   (e.g., *nifHDK* in an apparent operon) were retained. Each of the three nitrogenase subunits (i.e.,

578   Anf/Vnf/NifH, D, and K) were aligned individually with Clustal omega (123) specifying default

579   settings and the resultant alignment blocks were concatenated. The concatenated HDK sequences

580   were subjected to phylogenetic reconstruction as described above.

581

582   **Identification of homologs of Fd, Fld, and putative Fd- or Fld-reducing enzymes in**

583   **the genomes of diazotrophs.** To identify Fds and Flds in the genomes of putative diazotrophs,

584   we first identified all the Fd and Fld in representative organisms for each Nif sublineage. These

585   included *Azotobacter vinelandii*, *Klebsiella pneumoniae*, *Rhodopseudomonas palustris*, and

586   *Rhodobacter capsulatus* for Nif-A, *Clostridium pasteurianum* and *Methanosarcina barkeri* for

587   Nif-B, *Caldicellulosiruptor saccharolyticus* and *Methanocaldococcus vulcanius* for Nif-C, and

588   *Methanococcus maripaludis* for Nif-D. All protein sequences that were annotated as Fds or Flds

589   in the genomes of the above taxa were compiled. These protein sequences were subjected to

590   Conserved Domain Database (124) BLAST to identify the type of iron-sulfur clusters present in

591   the Fd and to confirm that the extracted Fld contained flavin binding motif(s). All the compiled

592   and curated Fds and Flds identified in these representative genomes were separately aligned

593   using Clustal Omega (123). The aligned Fd or Fld sequences were subjected to Maximum-

594   Likelihood phylogenetic reconstruction with RaxML (version 7.3.0) (125) specifying the LG

595   substitution matrix and the PROTGAMMA option to empirically cluster the sequences into

596   unique groups. In total 48 Fds (**Table 1**) and five unique Flds (**Table 2**) were identified among

597 diazotrophic genomes. The amino-acid sequences of all the Fds and Flds were then used as bait

598 sequences to identify homologs in the genomes of putative diazotrophs using BLASTp.

599

600   Putative enzymes that have been implicated in providing reductant to nitrogenase in the

601 form of reduced Fd and/or Fld include: *i*) pyruvate-flavodoxin oxidoreductase (PFOR) (36-39),

602 *ii*) ferredoxin-NADP$^+$ oxidoreductase (FNR) (36, 63), *iii*) *Rhodobacter* nitrogen fixation protein

603 (Rnf) (82), and *iv*) electron transfer flavoprotein (Fix) (28, 54, 55). Homologs of these enzymes

604 were compiled from genomes of putative diazotrophs using BLASTp. The amino-acid sequences

605 of PFOR from *Klebsiella pneumoniae* (WP_064371580) and *Rhodopseudomonas palustris*

606 (CAE30161), FNR from *Anabaena azollae* (WP_013193003) and from *Chlorobium tepidum*

607 TLS (AAM72739), Rnf from *Azotobacter vinelandii* (ACO81179- ACO81185), and Fix from

608 *Rhodopseudomonas palustris* (WP_011665892-WP_011665895) were used as query sequences.

609 In addition, [FeFe]-hydrogenase has been suggested to provide reductant for N_2 fixation in

610 *Clostridium pasteurianum* (40). Thus, the catalytic subunit of [FeFe] hydrogenase (HydA) from

611 *Chlamydomonas reinhardtii* (AAL23572) was used as query in a BLASTp search. Finally,

612 groups 3c, 3d, 4d and 4e [NiFe]-hydrogenases catalyze the reversible reduction of Fd with H_2

613 (42, 43, 103). A group 3c representative [NiFe]-hydrogenase from *Methanothermobacter*

614 *marburgensis* (WP_013296467) and a group 4 representative [NiFe]-hydrogenase from

615 *Pyrococcus abyssi* (WP_010867842.1) were used as BLASTp queries to identify all group 3 and

616 4 [NiFe]-hydrogenase large subunit sequences. Compiled sequences were aligned as described

617 above and manually curated to include only those enzymes that had proximal and distal Cys-Cys

618 pairs, the ligands for the [NiFe] active site (126, 127). Extracted [NiFe] homologous sequences

619 were then subjected to phylogenetic reconstruction, as described above, and assigned to their

27

620   respective groups based on phylogenetic coherence, as we have described previously (126, 127).

621   [NiFe]-hydrogenase homologs belonging to group 3c, 3d, 4d and 4e that have been proposed to

622   be capable of reducing Fd were only considered in this study (41-45).

623

624   **Identifying co-occurrence patterns in the distribution of Fds/Flds and putative**

625   **Fd/Fld-reducing enzymes.** A binary table was created based on the presence or absence of

626   identified Fds, Flds, and Fd- or Fld-reducing enzymes in diazotrophic genomes. To identify the

627   frequent patterns in this dataset, the Apriori algorithm was used as implemented with the 'arules'

628   package in R (128, 129). For simplicity, only the Fds, Flds, and Fd- or Fld-reducing enzymes

629   present in >20% of the genomes were considered for this analysis. We applied the Apriori

630   algorithm to our binary database while specifying the following parameters: A confidence

631   threshold of 0.60 and a support threshold of 0.2. The confidence threshold is the measure of the

632   statistical significance of an identified pattern within a dataset while the support threshold is

633   indicative of the relative abundance of a given pattern in a dataset. The Apriori algorithm works

634   in two phases with the first phase consisting of a scan of the entire database to objectively

635   identify and extract frequently observed proteins. In the second phase, the algorithm uses those

636   frequently occurring proteins to identify the dominant patterns in the distribution of proteins or

637   combinations of those proteins. In doing so, the algorithm identifies correlations (within the user

638   defined confidence and support threshold) between and among proteins in the database.

639

640   **Oxidation state of carbon.** The oxidation state of carbon for each amino-acid in each

641   protein sequence encoded in 359 putative diazotrophic genomes was calculated following the

642   algorithm mentioned in (113) using a custom python script (version 2.7.12). The calculated

643   oxidation number for each amino-acid was then normalized to the total number of amino acids

644   encoded in a genome to determine the average oxidation state of the proteome. This number was

645   determined for each genome that encoded a specific Nif isoform. These numbers were summed

646   and then divided by the total number of genomes to arrive at the average oxidation state of

647   carbon in a proteome inferred for a genome that encodes for a given nitrogenase lineage. A

648   Welch's t-test was conducted to determine the statistical significance of differences between the

649   average oxidation state of carbon in proteomes from different lineages using the stats package in

650   R (130). Box-plots were produced to show the distribution of the oxidation state of carbon in

651   proteomes inferred for genomes that encoded specific isoforms of nitrogenases using R (130).

652

653   **SUPPLEMENTAL MATERIAL**

654   **Supplemental Fig. 1**

655   **Supplemental Fig. 2**

656   **Supplemental Fig. 3**

657   **Supplemental Fig. 4**

658   **Supplemental Fig. 5**

659   **Supplemental Fig. 6**

660   **Supplemental Table 1**

661   **Supplemental Table 2**

662

663   **ACKNOWLEDGEMENTS**

664

REFERENCES

1. Raymond J, Siefert JL, Staples CR, Blankenship RE. 2004. The natural history of nitrogen fixation. Molecular Biology and Evolution 21:541-554.

2. Vitousek PM, Menge DNL, Reed SC, Cleveland CC. 2013. Biological nitrogen fixation: rates, patterns and ecological controls in terrestrial ecosystems. Philosophical Transactions of the Royal Society B-Biological Sciences 368.

3. Burk D, Lineweaver H, Horner CK. 1934. The specific influence of acidity on the mechanism of nitrogen fixation by *Azotobacter*. J Bact 27:325-340.

4. Falkowski PG. 1997. Evolution of the nitrogen cycle and its influence on the biological sequestration of CO_2 in the ocean. Nature 387:272-275.

5. Bulen W, LeComte J. 1966. The nitrogenase system from *Azotobacter*: two-enzyme requirement for N_2 reduction, ATP-dependent H_2 evolution, and ATP hydrolysis. Proc Natl Acad Sci U S A 56:979-986.

6. Peters JW, Boyd ES, Hamilton T, Rubio LM (ed). 2011. Biochemistry of Mo-Nitrogenase. Caister Academic Press, United Kingdom.

7. Boyd ES, Anbar AD, Miller S, Hamilton TL, Lavin M, Peters JW. 2011. A late methanogen origin for molybdenum-dependent nitrogenase. Geobiology 9:221-232.

8. Dos Santos PC, Fang Z, Mason SW, Setubal JC, Dixon R. 2012. Distribution of nitrogen fixation and nitrogenase-like sequences amongst microbial genomes. Bmc Genomics 13.

9. Raymond J, Siefert JL, Staples CR, Blankenship RE. 2004. The natural history of nitrogen fixation. Mol Biol Evol 21:541-554.

690   10.   McGlynn SE, Boyd ES, Peters JW, Orphan VJ. 2013. Classifying the metal dependence

691   of uncharacterized nitrogenases. Front Microbiol 3.

692   11.   Leigh JA. 2000. Nitrogen fixation in methanogens: The archaeal perspective. Curr Issues

693   Mol Biol 2:125-131.

694   12.   Mehta MP, Baross JA. 2006. Nitrogen fixation at 92 degrees C by a hydrothermal vent

695   archaeon. Science 314:1783-1786.

696   13.   Nishizawa M, Miyazaki J, Makabe A, Koba K, Takai K. 2014. Physiological and isotopic

697   characteristics of nitrogen fixation by hyperthermophilic methanogens: Key insights into

698   nitrogen anabolism of the microbial communities in Archean hydrothermal systems.

699   Geochim Cosmochim Acta 138:117-135.

700   14.   Boyd ES, Hamilton TL, Peters JW. 2011. An alternative path for the evolution of

701   biological nitrogen fixation. Front Microbiol 2.

702   15.   Joerger RD, Bishop PE, Bishop PE. 1988. Bacterial alternative nitrogen fixation systems.

703   Crit Rev Microbiol 16:1-14.

704   16.   Rubio LM, Ludden PW. 2008. Biosynthesis of the iron-molybdenum cofactor of

705   nitrogenase. Ann Rev Microbiol 62:93-111.

706   17.   Boyd ES, Peters JW. 2013. New insights into the evolutionary history of biological

707   nitrogen fixation. Front Microbiol 4.

708   18.   Boyd ES, Costas AMG, Hamilton TL, Mus F, Peters JW. 2015. Evolution of

709   molybdenum nitrogenase during the transition from anaerobic to aerobic Metabolism. J

710   Bact 197:1690-1699.

711 19. Stal LJ, Krumbein WE. 1985. Nitrogenase activity in the non-heterocystous

712   cyanobacterium *Oscillatoria* sp. grown under alternating light-dark cycles. Arch

713   Microbiol 143:67-71.

714 20. Fay P. 1992. Oxygen relations of nitrogen fixation in cyanobacteria. Microbiol Rev

715   56:340-373.

716 21. Poole RK, Hill S. 1997. Respiratory protection of nitrogenase activity in *Azotobacter*

717   *vinelandii* - Roles of the terminal oxidases. Biosci Rep 17:303-317.

718 22. Egener T, Martin DE, Sarkar A, Reinhold-Hurek B. 2001. Role of a ferredoxin gene

719   cotranscribed with the *nifHDK* operon in N_2 fixation and nitrogenase "switch-off" of

720   *Azoarcus* sp. Strain BH72. J Bact 183:3752-3760.

721 23. Sandmann G, Peleato ML, Fillat MF, Lazaro MC, Gomezmoreno C. 1990. Consequences

722   of the iron-dependent formation of ferredoxin and flavodoxin on photosynthesis and

723   nitrogen fixation on *Anabaena* Strains. Photosynth Res 26:119-125.

724 24. Yakunin AF, Gennaro G, Hallenbeck PC. 1993. Purification and properties of a *nif*-

725   specific flavodoxin from the photosynthetic bacterium *Rhodobacter capsulatus.* J Bact

726   175:6775-6780.

727 25. Martin AE, Burgess BK, Iismaa SE, Smartt CT, Jacobson MR, Dean DR. 1989.

728   Construction and characterization of an *Azotobacter vinelandii s*train with mutations in

729   the genes encoding flavodoxin and ferredoxin I. J Bact 171:3162-3167.

730 26. Yang Z-Y, Ledbetter R, Shaw S, Pence N, Tokmina-Lukaszewska M, Eilers B, Guo Q,

731   Pokhrel N, Cash VL, Dean DR. 2016. Evidence that the Pi release event is the rate-

732   limiting step in the nitrogenase catalytic cycle. Biochemistry 55:3625-3635.

733   27.   Gennaro G, Hubner P, Sandmeier U, Yakunin AF, Hallenbeck PC. 1996. Cloning,

734   characterization, and regulation of *nifF* from *Rhodobacter capsulatus*. J Bact 178:3949-

735   3952.

736   28.   Ledbetter RN, Garcia Costas AM, Lubner CE, Mulder DW, Tokmina-Lukaszewska M,

737   Artz JH, Patterson A, Magnuson TS, Jay ZJ, Duan HD, Miller J, Plunkett MH, Hoben JP,

738   Barney BM, Carlson RP, Miller AF, Bothner B, King PW, Peters JW, Seefeldt LC. 2017.

739   The electron bifurcating FixABCX protein complex from *Azotobacter vinelandii*:

740   Generation of low-potential reducing equivalents for nitrogenase catalysis. Biochemistry

741   56:4177-4190.

742   29.   Simpson FB, Burris RH. 1984. A nitrogen pressure of 50 atmospheres does not prevent

743   evolution of hydrogen by nitrogenase. Science 224:1095-1097.

744   30.   Tso M-YW, Ljones T, Burris R. 1972. Purification of the nitrogenase proteins from

745   *Clostridium pasteurianum*. Biochimica et Biophysica Acta (BBA)-Bioenergetics

746   267:600-604.

747   31.   Seefeldt LC, Hoffman BM, Dean DR. 2009. Mechanism of Mo-Dependent Nitrogenase.

748   Ann Rev Biochem 78:701-722.

749   32.   Bergström J, Eady RR, Thorneley RN. 1988. The vanadium-and molybdenum-containing

750   nitrogenases of *Azotobacter chroococcum*. Comparison of mid-point potentials and

751   kinetics of reduction by sodium dithionite of the iron proteins with bound magnesium

752   adenosine 5′-diphosphate. Biochem J 251:165-169.

753   33.   Dilworth MJ, Eldridge ME, Eady RR. 1992. Correction for creatine interference with the

754   direct indophenol measurement of NH_3 in steady-state nitrogenase assays. Anal Biochem

755   207:6-10.

756 34. Dilworth MJ, Eldridge ME, Eady RR. 1993. The molybdenum and vanadium

757   nitrogenases of *Azotobacter chroococcum*: effect of elevated temperature on N_2

758   reduction. Biochem J 289:395-400.

759 35. Eady RR. 1996. Structure−function relationships of alternative nitrogenases. Chem Rev

760   96:3013-3030.

761 36. Saeki K (ed). 2004. Electron transport to nitrogenase: Diverse routes for a common

762   destination. Kluwer, Dordrecht.

763 37. Nieva-Gomez D, Roberts GP, Klevickis S, Brill WJ. 1980. Electron transport to

764   nitrogenase in *Klebsiella pneumoniae*. Proc Natl Acad Sci U S A 77:2555-2558.

765 38. Wahl RC, Orme-Johnson WH. 1987. Clostridial pyruvate oxidoreductase and the

766   pyruvate-oxidizing enzyme specific to nitrogen fixation in *Klebsiella pneumoniae* are

767   similar enzymes. J Biol Chem 262:10489-10496.

768 39. Yakunin AF, Hallenbeck PC. 1998. Purification and characterization of pyruvate

769   oxidoreductase from the photosynthetic bacterium *Rhodobacter capsulatus*. Biochimica

770   Et Biophysica Acta-Bioenergetics 1409:39-49.

771 40. Therien J, Artz JH, Poudel S, Hamilton TL, Liu Z, Noone S, Adams M, King PW, Bryant

772   DA, Boyd ES, Peters JW. 2017. The physiological functions and structural determinants

773   of catalytic bias in the [FeFe]-hydrogenases CpI and CpII of *Clostridium pasteurianum*

774   strain W5. Front Microbiol 8:1305.

775 41. Bothe H, Schmitz O, Yates MG, Newton WE. 2010. Nitrogen fixation and hydrogen

776   metabolism in cyanobacteria. Microbiol Mol Biol Rev 74:529-551.

777 42. Greening C, Biswas A, Carere CR, Jackson CJ, Taylor MC, Stott MB, Cook GM,

778   Morales SE. 2016. Genomic and metagenomic surveys of hydrogenase distribution

779   indicate H$_2$ is a widely utilised energy source for microbial growth and survival. ISME J

780   10:761.

781   43.   Gutekunst K, Chen X, Schreiber K, Kaspar U, Makam S, Appel J. 2014. The

782   bidirectional NiFe-hydrogenase in *Synechocystis* sp PCC 6803 is reduced by flavodoxin

783   and ferredoxin and is essential under mixotrophic, nitrate-limiting conditions. J Biol

784   Chem 289:1930-1937.

785   44.   Khanna N, Lindblad P. 2015. Cyanobacterial hydrogenases and hydrogen metabolism

786   revisited: Recent progress and future prospects. Int J Mol Sci 16:10537-10561.

787   45.   Schmitz O, Boison G, Hilscher R, Hundeshagen B, Zimmer W, Lottspeich F, Bothe H.

788   1995. Molecular biological analysis of a bidirectional hydrogenase from cyanobacteria.

789   Eur J Biochem 233:266-276.

790   46.   Rey FE, Oda Y, Harwood CS. 2006. Regulation of uptake hydrogenase and effects of

791   hydrogen utilization on gene expression in *Rhodopseudomonas palustris*. J Bact

792   188:6143-6152.

793   47.   Herrmann G, Jayamani E, Mai G, Buckel W. 2008. Energy conservation via electron-

794   transferring flavoprotein in anaerobic bacteria. J Bact 190:784-791.

795   48.   Anraku Y. 1988. Bacterial electron-transport chains. Ann Rev Biochem 57:101-132.

796   49.   Watt G, Burns A, Lough S, Tennent D. 1980. Redox and spectroscopic properties of

797   oxidized MoFe protein from *Azotobacter vinelandii*. Biochemistry 19:4926-4932.

798   50.   Lanzilotta WN, Ryle MJ, Seefeldt LC. 1995. Nucleotide hydrolysis and protein

799   conformational changes in *Azotobacter vinelandii* nitrogenase iron protein: Defining the

800   function of aspartate 129. Biochemistry 34:10713-10723.

36

801   51.   Braaksma A, Haaker H, Grande HJ, Veeger C. 1982. The effect of the redox potential on

802   the activity of the nitrogenase and on the Fe-protein of *Azotobacter vinelandi*. FEBS J

803   121:483-491.

804   52.   Dutton P. 1986. Energy transduction in anoxygenic photosynthesis, p 197-237,

805   Photosynthesis III. Springer.

806   53.   Jagannathan B, Golbeck J (ed). 2009. Photosynthesis: microbial. Oxford,

807   54.   Earl C, Ronson C, Ausubel F. 1987. Genetic and structural analysis of the *Rhizobium*

808   *meliloti fixA*, *fixB*, *fixC*, and *fixX* genes. J Bact 169:1127-1136.

809   55.   Edgren T, Nordlund S. 2004. The *fixABCX* genes in *Rhodospirillum rubrum* encode a

810   putative membrane complex participating in electron transfer to nitrogenase. J Bact

811   186:2052-2060.

812   56.   Biegel E, Schmidt S, Gonzalez JM, Muller V. 2011. Biochemistry, evolution and

813   physiological function of the Rnf complex, a novel ion-motive electron transport complex

814   in prokaryotes. Cell Mol Life Sci 68:613-634.

815   57.   Hess V, Schuchmann K, Muller V. 2013. The ferredoxin: NAD(+) oxidoreductase (Rnf)

816   from the acetogen *Acetobacterium woodii* requires Na^+ and is reversibly coupled to the

817   membrane potential. J Biol Chem 288:31496-31502.

818   58.   Jouanneau Y, Jeong HS, Hugo N, Meyer C, Willison JC. 1998. Overexpression in

819   *Escherichia coli* of the *rnf* genes from *Rhodobacter capsulatus* - Characterization of two

820   membrane-bound iron-sulfur proteins. Eur J Biochem 251:54-64.

821   59.   Sarkar A, Köhler J, Hurek T, Reinhold-Hurek B. 2012. A novel regulatory role of the Rnf

822   complex of *Azoarcus sp*. strain BH72. Mol Microbiol 83:408-422.

823   60.   Schrautemeier B, Böhme H. 1985. A distinct ferredoxin for nitrogen fixation isolated

824   from heterocysts of the cyanobacterium *Anabaena variabilis*. FEBS Lett 184:304-308.

825   61.   Apte S, Rowell P, Stewart W. 1978. Electron donation to ferredoxin in heterocysts of the

826   N_2-fixing alga *Anabaena cylindrica*. Proc Royal Soc B 200:1-25.

827   62.   Aliverti A, Faber R, Finnerty CM, Ferioli C, Pandini V, Negri A, Karplus PA, Zanetti G.

828   2001. Biochemical and Crystallographic Characterization of Ferredoxin-NADP(+)

829   Reductase from Nonphotosynthetic Tissues. Biochemistry 40:14501-14508.

830   63.   Isas JM, Yannone SM, Burgess BK. 1995. *Azotobacter vinelandii* NADPH-rerredoxin

831   reductase cloning, sequencing, and overexpression. J Biol Chem 270:21258-21263.

832   64.   Sattley WM, Madigan MT, Swingley WD, Cheung PC, Clocksin KM, Conrad AL,

833   Dejesa LC, Honchak BM, Jung DO, Karbach LE, Kurdoglu A, Lahiri S, Mastrian SD,

834   Page LE, Taylor HL, Wang ZT, Raymond J, Chen M, Blankenship RE, Touchman JW.

835   2008. The genome of *Heliobacterium modesticaldum*, a phototrophic representative of

836   the Firmicutes containing the simplest photosynthetic apparatus. J Bact 190:4687-4696.

837   65.   Seo D, Sakurai H. 2002. Purification and characterization of ferredoxin-NAD(P)(+)

838   reductase from the green sulfur bacterium *Chlorobium tepidum*. Biochim Biophys Acta

839   1597:123-132.

840   66.   Muraki N, Seo D, Shiba T, Sakurai T, Kurisu G. 2010. Asymmetric dimeric structure of

841   ferredoxin-NAD(P)$^+$ oxidoreductase from the green sulfur bacterium *Chlorobaculum*

842   *tepidum*: implications for binding ferredoxin and NADP$^+$. J Mol Biol 401:403-14.

843   67.   Petering D, Fee JA, Palmer G. 1971. The oxygen sensitivity of spinach ferredoxin and

844   other iron-sulfur proteins. The formation of protein-bound sulfur-zero. J Biol Chem

845   246:643-653.

846 68. Knight E, Hardy R. 1966. Isolation and characteristics of flavodoxin from nitrogen-fixing

847   *Clostridium pasteurianum*. J Biol Chem 241:2752-2756.

848 69. Lovenberg W, Buchanan BB, Rabinowitz JC. 1963. Studies on the chemical nature of

849   clostridial ferredoxin. J Biol Chem 238:3899-3913.

850 70. Imlay JA. 2006. Iron-sulphur clusters and the problem with oxygen. Mol Microbiol

851   59:1073-1082.

852 71. Fillat MF, Sandmann G, Gomez-Moreno C. 1988. Flavodoxin from the nitrogen-fixing

853   cyanobacterium *Anabaena* PCC 7119. Arch Microbiol 150:160-164.

854 72. Lodeyro AF, Ceccoli RD, Karlusich JJP, Carrillo N. 2012. The importance of flavodoxin

855   for environmental stress tolerance in photosynthetic microorganisms and transgenic

856   plants. Mechanism, evolution and biotechnological potential. FEBS Lett 586:2917-2924.

857 73. Bennett LT, Jacobson MR, Dean DR. 1988. Isolation, sequencing, and mutagenesis of the

858   *nifF* gene encoding flavodoxin from *Azotobacter vinelandii*. J Biol Chem 263:1364-1369.

859 74. Hill S, Kavanagh EP. 1980. Roles of *nifF* and *nifJ* gene products in electron transport to

860   nitrogenase in *Klebsiella pneumoniae*. J Bact 141:470-475.

861 75. Shah VK, Stacey G, Brill WJ. 1983. Electron transport to nitrogenase - Purification and

862   characterization of pyruvate, flavodoxin oxidoreductase, the *nifJ*-gene product. J Biol

863   Chem 258:2064-2068.

864 76. Yoch DC, Arnon DI. 1972. Two biologically active ferredoxins from the aerobic

865   nitrogen-fixing bacterium, *Azotobacter vinelandii*. J Biol Chem 247:4514-4520.

866 77. Jouanneau Y, Meyer C, Naud I, Klipp W. 1995. Characterization of an *fdxN* mutant of

867   *Rhodobacter capsulatus i*ndicates that Ferredoxin-I serves as electron donor to

868   nitrogenase. Biochim Biophys Acta-Bioenerg 1232:33-42.

869   78.   McRose DL, Zhang XN, Kraepiel AML, Morel FMM. 2017. Diversity and activity of

870   alternative nitrogenases in sequenced genomes and coastal environments. Front

871   Microbiol 8.

872   79.   Dekas AE, Poretsky RS, Orphan VJ. 2009. Deep-sea Archaea fix and share nitrogen in

873   methane-consuming microbial consortia. Science 326:422-426.

874   80.   Bolhuis H, Severin I, Confurius-Guns V, Wollenzien UIA, Stal LJ. 2010. Horizontal

875   transfer of the nitrogen fixation gene cluster in the cyanobacterium *Microcoleus*

876   *chthonoplastes*. ISME J 4:121-130.

877   81.   Armengaud J, Meyer C, Jouanneau Y. 1994. Recombinant expression of the *fdxD* gene of

878   *Rhodobacter capsulatus* and characterization of its product, a [2Fe-2S] ferredoxin.

879   Biochem J 300:413-418.

880   82.   Schmehl M, Jahn A, Vilsendorf AMZ, Hennecke S, Masepohl B, Schuppler M, Marxer

881   M, Oelze J, Klipp W. 1993. Identification of a new class of nitrogen-fixation genes in

882   *Rhodobacter capsulatus* - A putative membrane complex involved in electron transport to

883   nitrogenase. Mol Gen Genet 241:602-615.

884   83.   Hamilton TL, Ludwig M, Dixon R, Boyd ES, Dos Santos PC, Setubal JC, Bryant DA,

885   Dean DR, Peters JW. 2011. Transcriptional profiling of nitrogen fixation in *Azotobacter*

886   *vinelandii*. J Bact 193:4477-4486.

887   84.   Jouanneau Y, Meyer C, Gaillard J, Forest E, Gagnon J. 1993. Purification and

888   characterization of a novel dimeric ferredoxin (FdIII) from *Rhodobacter capsulatus*. J

889   Biol Chem 268:10636-10644.

890   85.   Moreno-Vivian C, Hennecke S, Pühler A, Klipp W. 1989. Open reading frame 5 (ORF5),

891   encoding a ferredoxin like protein, and *nifQ* are cotranscribed with *nifE*, *nifN*, *nifX*, and

892   ORF4 in *Rhodobacter capsulatus.* J Bact 171:2591-2598.

893   86.   Masepohl B, Schölisch K, Görlitz K, Kutzki C, Böhme H. 1997. The heterocyst-specific

894   *fdxH* gene product of the cyanobacterium *Anabaena* sp. PCC 7120 is important but not

895   essential for nitrogen fixation. Mol Gen Genet 253:770-776.

896   87.   Souza ALF, Invitti AL, Rego FGM, Monteiro RA, Klassen G, Souza EM, Chubatsu LS,

897   Pedrosa FO, Rigo LU. 2010. The involvement of the *nif*-associated ferredoxin-like genes

898   *fdxA* and *fdxN* of *Herbaspirillum seropedicae* in nitrogen fixation. J Microbiol 48:77-83.

899   88.   Hageman RV, Burris RH. 1978. Nitrogenase and nitrogenase reductase associate and

900   dissociate with each catalytic cycle. Proc Natl Acad Sci U S A 75:2699-2702.

901   89.   Malkin R, Aparicio PJ, Arnon DI. 1974. The isolation and characterization of a new iron-

902   sulfur protein from photosynthetic membranes. Proc Natl Acad Sci USA 71:2362-2366.

903   90.   Hallenbeck PC, Gennaro G. 1998. Stopped-flow kinetic studies of low potential electron

904   carriers of the photosynthetic bacterium, *Rhodobacter capsulatus*: ferredoxin I and NifF.

905   Biochim Biophys Acta-Bioenerg 1365:435-442.

906   91.   Hallenbeck PC, Jouanneau Y, Vignais PM. 1982. Purification and molecular properties

907   of a soluble ferredoxin from *Rhodopseudomonas capsulata*. Biochim Biophys Acta-

908   Bioenerg 681:168-176.

909   92.   Edgren T, Nordlund S. 2005. Electron transport to nitrogenase in *Rhodospirillum rubrum*:

910   Identification of a new *fdxN* gene encoding the primary electron donor to nitrogenase.

911   FEMS Microbiol Lett 245:345-351.

41

912   93.   Klipp W, Reilander H, Schluter A, Krey R, Puhler A. 1989. The *Rhizobium meliloti fdxN*

913   gene encoding a ferredoxin-like protein is necessary for nitrogen fixation and is

914   cotranscribed with *nifA* and *nifB*. Mol Gen Genet 216:293-302.

915   94.   Armengaud J, Meyer C, Jouanneau Y. 1997. A [2Fe-2S] ferredoxin (FdVI) is essential

916   for growth of the photosynthetic bacterium *Rhodobacter capsulatus*. J Bact 179:3304-

917   3309.

918   95.   Schrautemeier B, Cassing A, Bohme H. 1994. Characterization of the genome region

919   encoding an FdxH-type ferredoxin and a new 2[4fe-4s] ferredoxin from the

920   nonheterocystous, nitrogen-fixing cyanobacterium *Plectonema boryanum* PCC-73110. J

921   Bact 176:1037-1046.

922   96.   Ermakova M, Battchikova N, Richaud P, Leino H, Kosourov S, Isojarvi J, Peltier G,

923   Flores E, Cournac L, Allahverdiyeva Y, Aro EM. 2014. Heterocyst-specific flavodiiron

924   protein Flv3B enables oxic diazotrophic growth of the filamentous cyanobacterium

925   *Anabaena* sp. PCC 7120. Proc Natl Acad Sci U S A 111:11205-11210.

926   97.   Peden EA, Boehm M, Mulder DW, Davis R, Old WM, King PW, Ghirardi ML, Dubini

927   A. 2013. Identification of global ferredoxin interaction networks in *Chlamydomonas*

928   *reinhardtii*. J Biol Chem 288:35192-35209.

929   98.   Grabau C, Schatt E, Jouanneau Y, Vignais PM. 1991. A new [2Fe-2S] ferredoxin from

930   *Rhodobacter capsulatus*. Coexpression with a 2 [4Fe-4S] ferredoxin in *Escherichia coli*.

931   J Biol Chem 266:3294-3299.

932   99.   Deistung J, Cannon FC, Cannon MC, Hill S, Thorneley RNF. 1985. Electron transfer to

933   nitrogenase in *Klebsiella pneumoniae* - *nifF g*ene cloned and the gene product, a

934   flavodoxin, purified. Biochem J 231:743-753.

935   100.   Drummond MH. 1985. The base sequence of the *nifF g*ene of *Klebsiella pneumoniae* and

936   homology of the predicted amino acid sequence of its protein product to other

937   flavodoxins. Biochem J 232:891-896.

938   101.   Roberts GP, MacNeil T, MacNeil D, Brill WJ. 1978. Regulation and characterization of

939   protein products coded by the *nif* (nitrogen fixation) genes of *Klebsiella pneumoniae*. J

940   Bact 136:267-279.

941   102.   Arnold W, Rump A, Klipp W, Priefer UB, Puhler A. 1988. Nucleotide sequence of a

942   24,206-base-pair DNA fragment carrying the entire nitrogen fixation gene cluster of

943   *Klebsiella pneumoniae*. J Mol Biol 203:715-738.

944   103.   Peters JW, Schut GJ, Boyd ES, Mulder DW, Shepard EM, Broderick JB, King PW,

945   Adams MWW. 2015. FeFe- and NiFe-hydrogenase diversity, mechanism, and

946   maturation. Biochim Biophys Acta-Mol Cell Res 1853:1350-1369.

947   104.   Poudel S, Tokmina-Lukaszewska M, Colman DR, Refai M, Schut GJ, King PW, Maness

948   PC, Adams MWW, Peters JW, Bothner B, Boyd ES. 2016. Unification of [FeFe]-

949   hydrogenases into three structural and functional groups. Biochim Biophys Acta-Gen

950   Subj 1860:1910-1921.

951   105.   Kaminski PA, Norel F, Desnoues N, Kush A, Salzano G, Elmerich C. 1988.

952   Characterization of the *fixABC* region of *Azorhizobium caulinodans* ors571 and

953   identification of a new nitrogen fixation gene. Mol Gen Genet 214:496-502.

954   106.   Gubler M, Hennecke H. 1986. *fixA*, *B* and *C* genes are essential for symbiotic and free-

955   living, microaerobic nitrogen fixation. FEBS Lett 200:186-192.

956   107.   Garcia Costas AM, Poudel S, Miller A-F, Schut GJ, Ledbetter RN, Fixen KR, Seefeldt

957   LC, Adams MW, Harwood CS, Boyd ES. 2017. Defining electron bifurcation in the

958   electron transferring flavoprotein family. J Bact:JB. 00440-17.

959   108.   Furdui C, Ragsdale SW. 2000. The role of pyruvate ferredoxin oxidoreductase in

960   pyruvate synthesis during autotrophic growth by the Wood-Ljungdahl pathway. J Biol

961   Chem 275:28494-28499.

962   109.   Neuer G, Bothe H. 1985. Electron donation to nitrogenase in heterocysts of

963   cyanobacteria. Arch Microbiol 143:185-191.

964   110.   Lubitz W, Ogata H, Rudiger O, Reijerse E. 2014. Hydrogenases. Chem Rev 114:4081-

965   4148.

966   111.   Volbeda A, Charon MH, Piras C, Hatchikian EC, Frey M, Fontecilla-Camps JC. 1995.

967   Crystal structure of the nickel-iron hydrogenase from Desulfovibrio gigas. Nature

968   373:580-7.

969   112.   Shock EL, Boyd ES. 2015. Principles of Geobiochemistry. Elements 11:395-401.

970   113.   Dick JM, Shock EL. 2011. Calculation of the relative chemical stabilities of proteins as a

971   function of temperature and redox chemistry in a hot spring. PLOS ONE 6.

972   114.   Dick JM. 2014. Average oxidation state of carbon in proteins. J Roy Soc Interface 11.

973   115.   Singh BB, Curdt I, Jakobs C, Schomburg D, Bisen PS, Bohme H. 1999. Identification of

974   amino acids responsible for the oxygen sensitivity of ferredoxins from *Anabaena*

975   *variabilis* using site-directed mutagenesis. Biochim Biophys Acta-Bioenerg 1412:288-

976   294.

977   116.   Jannasch HW, Mottl MJ. 1985. Geomicrobiology of Deep-Sea Hydrothermal Vents.

978   Science 229:717-725.

44

979   117.   Anantharaman K, Breier JA, Sheik CS, Dick GJ. 2013. Evidence for hydrogen oxidation

980   and metabolic plasticity in widespread deep-sea sulfur-oxidizing bacteria. Proc Natl Acad

981   Sci USA 110:330-335.

982   118.   Zheng Y, Harris DF, Yu Z, Fu Y, Poudel S, Ledbetter RN, Fixen KR, Yang Z-Y, Boyd

983   ES, Lidstrom ME. 2018. A pathway for biological methane production using bacterial

984   iron-only nitrogenase. Nature Micro 1.

985   119.   Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li WZ, Lopez R, McWilliam H,

986   Remmert M, Soding J, Thompson JD, Higgins DG. 2011. Fast, scalable generation of

987   high-quality protein multiple sequence alignments using Clustal Omega. Molecular

988   Systems Biology 7.

989   120.   Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis

990   of large phylogenies. Bioinformatics 30:1312-1313.

991   121.   Boyd ES, Hamilton TL, Peters JW. 2011. An alternative path for the evolution of

992   biological nitrogen fixation. Frontiers in Microbiology 2.

993   122.   Boyd ES, Peters JW. 2013. New insights into the evolutionary history of biological

994   nitrogen fixation. Frontiers in Microbiology 4.

995   123.   Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li WZ, Lopez R, McWilliam H,

996   Remmert M, Soding J, Thompson JD, Higgins DG. 2011. Fast, scalable generation of

997   high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7.

998   124.   Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu SN, Chitsaz F, Geer LY, Geer RC,

999   He J, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Marchler GH, Song JS, Thanki N,

1000   Wang ZX, Yamashita RA, Zhang DC, Zheng CJ, Bryant SH. 2015. CDD: NCBI's

1001   conserved domain database. Nucl Acids Res 43:D222-D226.

1002   125.   Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis

1003   of large phylogenies. Bioinformatics 30:1312-3.

1004   126.   Boyd ES, Schut GJ, Adams MWW, Peters JW. 2014. Hydrogen metabolism and the

1005   evolution of biological respiration. Microbe 9:361-367.

1006   127.   Schut GJ, Zadvornyy O, Wu C-H, Peters JW, Boyd ES, Adams MWW. 2016. The role of

1007   geochemistry and energetics in the evolution of modern respiratory complexes from a

1008   proton-reducing ancestor. Biochim Biophys Acta - Bioenerg 1857:958-970.

1009   128.   Hahsler M, Chelluboina S, Hornik K, Buchta C. 2011. The arules R-package ecosystem:

1010   Analyzing interesting patterns from large transaction data sets. J Mach Learning Res

1011   12:2021-2025.

1012   129.   Agrawal R, Srikant R. Fast algorithms for mining association rules, p 487-499. *In* (ed),

1013   130.   Team RDC. 2010. R: A language and environment for statistical computing. R

1014   Foundation for Statistical Computing, Vienna, Austria. ISBN 3–900051–07–0, URL:

1015   http://www. R-project. org,

1016   131.   Kaneko T, Nakamura Y, Wolk CP, Kuritz T, Sasamoto S, Watanabe A, Iriguchi M,

1017   Ishikawa A, Kawashima K, Kimura T, Kishida Y, Kohara M, Matsumoto M, Matsuno A,

1018   Muraki A, Nakazaki N, Shimpo S, Sugimoto M, Takazawa M, Yamada M, Yasuda M,

1019   Tabata S. 2001. Complete genomic sequence of the filamentous nitrogen-fixing

1020   cyanobacterium *Anabaena* sp. strain PCC 7120. DNA Res 8:205-13; 227-53.

1021   132.   Jacobson BL, Chae YK, Böhme H, Markley JL, Holden HM. 1992. Crystallization and

1022   preliminary analysis of oxidized, recombinant, heterocyst [2Fe-2S] ferredoxin from

1023   *Anabaena* 7120. Archives of biochemistry and biophysics 294:279-281.

1024   133.   Rypniewski WR, Breiter DR, Benning MM, Wesenberg G, Oh BH, Markley JL,

1025   Rayment I, Holden HM. 1991. Crystallization and structure determination of 2.5 Å

1026   resolution of the oxidized [2Fe-2S] ferredoxin isolated from *Anabaena* 7120.

1027   Biochemistry 30:4126-4131.

1028   134.   Jacobson M, Brigle K, Bennett L, Setterquist R, Wilson M, Cash V, Beynon J, Newton

1029   W, Dean D. 1989. Physical and genetic map of the major nif gene cluster from

1030   *Azotobacter vinelandii*. J Bact 171:1017-1027.

1031   135.   Setubal JC, dos Santos P, Goldman BS, Ertesvag H, Espin G, Rubio LM, Valla S,

1032   Almeida NF, Balasubramanian D, Cromes L, Curatti L, Du ZJ, Godsy E, Goodner B,

1033   Hellner-Burris K, Hernandez JA, Houmiel K, Imperial J, Kennedy C, Larson TJ, Latreille

1034   P, Ligon LS, Lu J, Maerk M, Miller NM, Norton S, O'Carroll IP, Paulsen I, Raulfs EC,

1035   Roemer R, Rosser J, Segura D, Slater S, Stricklin SL, Studholme DJ, Sun J, Viana CJ,

1036   Wallin E, Wang BM, Wheeler C, Zhu HJ, Dean DR, Dixon R, Wood D. 2009. Genome

1037   sequence of *Azotobacter vinelandii*, an obligate aerobe specialized to support diverse

1038   anaerobic metabolic processes. J Bact 191:4534-4545.

1039   136.   Schatt E, Jouanneau Y, Vignais P. 1989. Molecular cloning and sequence analysis of the

1040   structural gene of ferredoxin I from the photosynthetic bacterium *Rhodobacter*

1041   *capsulatus*. J Bact 171:6218-6226.

1042   137.   Reyntjens B, Jollie DR, Stephens PJ, GaoSheridan HS, Burgess BK. 1997. Purification

1043   and characterization of a *fixABCX*-linked 2[4Fe-4S] ferredoxin from *Azotobacter*

1044   *vinelandii.* J Biol Inorg Chem 2:595-602.

1045   138.   Saeki K, Suetsugu Y, Tokuda K-I, Miyatake Y, Young D, Marrs B, Matsubara H. 1991.

1046   Genetic analysis of functional differences among distinct ferredoxins in *Rhodobacter*

1047   *capsulatus*. J Biol Chem 266:12889-12895.

1048   139.   Larimer FW, Chain P, Hauser L, Lamerdin J, Malfatti S, Do L, Land ML, Pelletier DA,

1049   Beatty JT, Lang AS, Tabita FR, Gibson JL, Hanson TE, Bobst C, Torres JLTY, Peres C,

1050   Harrison FH, Gibson J, Harwood CS. 2004. Complete genome sequence of the

1051   metabolically versatile photosynthetic bacterium *Rhodopseudomonas palustris*. Nature

1052   Biotech 22:55-61.

1053

1054   **FIGURE LEGENDS**

1055

1056   **Figure 1.** Phylogenetic reconstruction of a concatenation of H, D, and K subunits of nitrogenase

1057   and uncharacterized nitrogenase paralogs (n= 420 concatenated protein sequences). All nodes

1058   shown exhibited bootstraps supports of >90% (out of 1000 bootstrap replicates) except where

1059   black boxes (>70%) are shown. Abbreviations: Nif, molybdenum (Mo) nitrogenase; Anf, iron

1060   only (Fe) nitrogenase; Vnf, vanadium (V) nitrogenase; Unc, uncharacterized nitrogenase-like

1061   proteins.

1062

1063   **Figure 2.** Homologs of nitrogenase identified in the genomes of putative diazotrophs. A)

1064   Histogram depicting the percentage of diazotrophs within each specified Nif lineage (see **Fig. 1**)

1065   that are aerobic, anaerobic, or facultatively anaerobic, as determined from literature surveys of

1066   cultivated organisms. B) Venn diagram representing the number of genomes that encode for one

1067   or more specified lineages of nitrogenase (see **Fig. 1**). Genomes that encode for Nif-C and Nif-D

1068   do not encode for other isoforms of nitrogenase and are thus depicted as separate.

1069

1070   **Figure 3**. Histogram depicting the percentage of enzyme homologs putatively involved in Fd/Fld

1071   reduction identified in the genomes comprising a specified group. Aerobic/facultative anaerobic

1072   and anaerobic organisms are further classified as phototrophs or chemotrophs. Phototrophs in

1073   aerobes/facultative anaerobes are further classified based on whether they are oxygenic or

1074   anoxygenic phototrophs while anaerobic phototrophs are classified as either purple sulfur/non-

1075   sulfur bacteria or green sulfur/non-sulfur bacteria. Importantly, it is not clear from literature

1076   surveys that anaerobic purple bacteria strains [denoted by an asterisk (*)] were robustly tested for

1077   their ability to use O_2. Abbreviations: FNR, ferredoxin-NADP$^+$ oxidoreductase; PFOR, pyruvate-

1078   flavodoxin oxidoreductase; Rnf, *Rhodobacter* nitrogen fixation protein; FeFe, iron only

1079   hydrogenase; NiFe, nickel-iron hydrogenase; FixABCX, electron transfer flavoprotein involved

1080   in nitrogen fixation.

1081

1082   **Figure 4.** Bubble plot depicting the dominant patterns in the distribution of putative electron

1083   carrier protein homologs (Fds/Flds), enzyme homologs that putatively reduce Fd/Fld, and

1084   specified Nif lineages, as determined by the Apriori algorithm (129). Each unique pattern is

1085   given as a bubble and the color represents the confidence value or the statistical significance in

1086   the co-distribution of specified proteins (only confidence values of ≥ 0.6 are presented). The size

1087   of the bubble represents the support value (>=0.2) or the frequency that two proteins are

1088   identified in the same genome. For simplicity, only the proteins that were present in >20% of the

1089   diazotrophic genomes for each specified nitrogenase lineage were considered in this analysis.

1090   Abbreviations: PFOR, pyruvate-Fld oxidoreductase; Rnf, *Rhodobacter* nitrogen fixation protein;

1091   FeFe, iron-only hydrogenase; NiFe, nickel-iron hydrogenase; FixABCX, electron transfer

1092   flavoprotein that is involved in nitrogen fixation. Abbreviations for Fds and Flds are presented in

1093   **Tables 1** and **2**, respectively.

1094

1095   **Figure 5**. Box plot representing the average oxidation state of proteomes in each genome

1096   encoding the specified nitrogenase isoform. Here, the box represents the interquartile range with

1097   the whiskers that show the full range of the data. Outlier values are represented as circles and the

1098   horizontal black bold line within the box represents the median value.

1099

1100

1101

1102

1103

1104

1105

1106

1107

1108 **Table 1.** Ferredoxin (Fd) homologs identified in the genomes of putative diazotrophs. The protein annotation, sequence identifier for
1109 a representative protein homolog, inferred type of [FeS] cluster in a given protein homolog (as predicted via the conserved domain
1110 database (124)), and a literature reference is provided for each Fd identified, if a reference is available.

Annotation	Protein Id	Type of cluster	Source	Annotation	Protein Id	Type of cluster	Source
asl2914	WP_013190310	2[4Fe-4S]	(131)	KpFd3	BAS33241	[2Fe-2S]	This study
FdxH	WP_013190616	[2Fe-2S]	(132)	KpFd4	BAS35880	CxxCxxCC	This study
PetF	WP_013189829	[2Fe-2S]	(133)	MvFd1	ACX73019	2[4Fe-4S]	This study
FdxA	ACO80005	[3Fe-4S] [4Fe-4S]	(134)	MvFd2	ACX72493	2[4Fe-4S]	This study
FdxI	ACO76607	[4Fe-4S]	(135)	MvFd3	ACX73480	2[4Fe-4S]	This study
FdxN	ACO81189	2[4Fe-4S]	(83, 136)	MvFd4	ACX72328	2[4Fe-4S]	This study
VnfF	ACO76526	[4Fe-4S]	(137)	MvFd5	ACX73314	2[4Fe-4S]	This study
XylT	ACO77112	CxxxxCxxxxCxxC	(135)	MvFd6	ACX73502	2[4Fe-4S]	This study
AvFd1	AGK16019	[2Fe-2S]	This study	MvFd7	ACX72560	2[4Fe-4S]	This study
AvFd2	AGK16622	[2Fe-2S]	This study	MvFd8	ACX72349	2[4Fe-4S]	This study
AvFd3	AGK13276	2[4Fe-4S]	This study	MmFd1	CAF29654	2[4Fe-4S]	This study
CsFd1	ABP66178	2[4Fe-4S]	This study	MbFd1	AKB59153	2[4Fe-4S]	This study
CsFd2	ABP66040	2[4Fe-4S]	This study	MbFd2	AKB57181	2[4Fe-4S]	This study
CsFd3	ABP66582	2[4Fe-4S]	This study	MbFd3	AKB57361	2[4Fe-4S]	This study
CsFd4	ABP67884	2[4Fe-4S]	This study	MbFd4	AKB58903	2[4Fe-4S]	This study
CsFd5	ABP67132	2[4Fe-4S]	This study	MbFd5	AKB59009	2[4Fe-4S]	This study
CsFd6	ABP67141	2[4Fe-4S]	This study	FdxC	ADE87009	[2Fe-2S]	(98, 138)
CpFd1	AJA47502	[2Fe-2S]	This study	FdxD	ADE84338	[2Fe-2S]	(81)
CpFd2	AJA47129	2[4Fe-4S]	This study	FdxE	ADE86134	[2Fe-2S]	(94)
CpFd3	AJA49513	2[4Fe-4S]	This study	FdxB	ABJ08445	2[4Fe-4S]	(82)
CpFd4	AJA46278	[4Fe-4S]	This study	Fer1	ADU46354	[4Fe-4S]	(139)
CpFd5	AJA49585	2[4Fe-4S]	This study	FerN	ACF03599	[4Fe-4S]	(139)
KpFd1	BAS34286	[2Fe-2S]	This study	RpFd1	ABJ04561	2[4Fe-4S]	This study
KpFd2	BAS37351	[2Fe-2S]	This study				

1111   **Table 2.** Flavodoxin (Fld) homologs identified in the genomes of putative diazotrophs. The
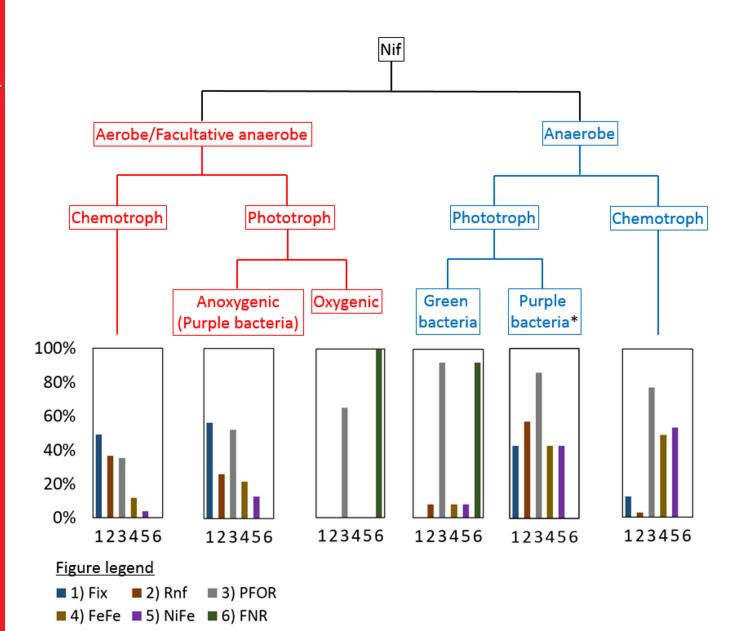1112   protein annotation, sequence identifier for a representative protein homolog, and a literature
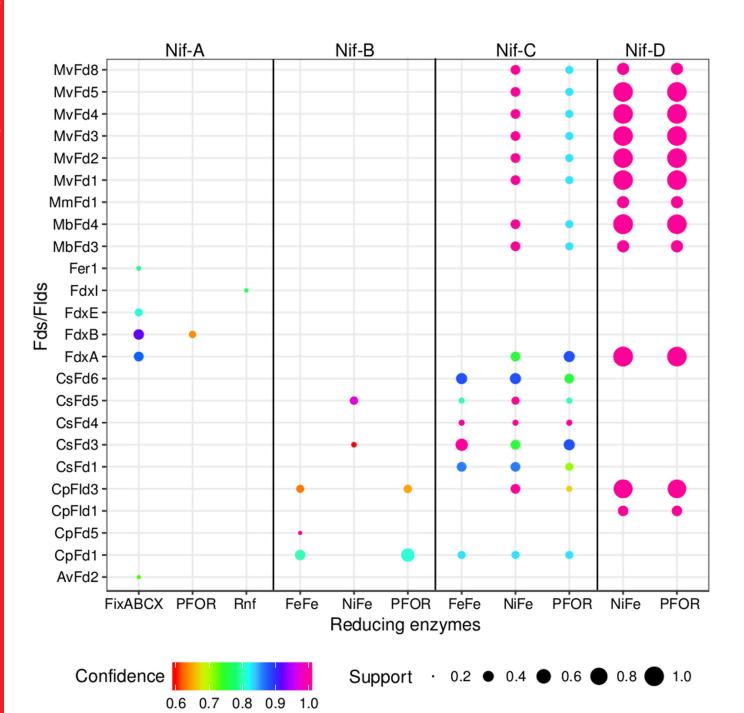1113   reference is provided for each Fld identified, if a reference is available.

Annotation	Accession Number	Source
NifF	ACO76434	(37)
CpFld1	AJA46461	This study
CpFld2	AJA47463	This study
CpFld3	AJA47660	This study
FldA	WP_011157672	(94)

1114

Nif-B

Anf

Vnf

Nif-A

Nif-D

Unc

Nif-C

0.4

A)

B)

Figure legend

- 1) Fix
- 2) Rnf
- 3) PFOR
- 4) FeFe
- 5) NiFe
- 6) FNR