Reconstructed ancient nitrogenases suggest Mo-specific ancestry

Amanda K. Garcia[1], Hanon McShea[2,a], Bryan Kolaczkowski[3], Betül Kaçar[1,4]

[1]Department of Molecular and Cellular Biology, University of Arizona, Tucson, Arizona

[2]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts

[3]Department of Microbiology and Cell Science, University of Florida, Gainesville, Florida

[4]Department of Astronomy and Steward Observatory, University of Arizona, Tucson, Arizona

[a]Present address: Department of Earth System Science, Stanford University, Stanford, California

Corresponding author information:

Betul Kacar

1007 E. Lowell St., Tucson, AZ 85721

betul@arizona.edu

ABSTRACT

The nitrogenase metalloenzyme family, essential for supplying fixed nitrogen to the biosphere, is one of life's key biogeochemical innovations. The three isozymes of nitrogenase differ in their metal dependence, each binding either a FeMo-, FeV-, or FeFe-cofactor for the reduction of nitrogen. The history of nitrogenase metal dependence has been of particular interest due to the possible implication that ancient marine metal availabilities have significantly constrained nitrogenase evolution over geologic time. Here, we combine phylogenetics and ancestral sequence reconstruction — a method by which inferred, historical protein sequence information can be linked to functional molecular properties — to reconstruct the metal dependence of ancient nitrogenases. Inferred ancestral nitrogenase sequences at the deepest nodes of the phylogeny

suggest that ancient nitrogenases were Mo-dependent. We find that active-site sequence identity can reliably distinguish extant Mo-nitrogenases from V- and Fe-nitrogenases, as opposed to modeled active-site structural features that cannot be used to reliably classify nitrogenases of unknown metal dependence. Taxa represented by early-branching nitrogenase lineages lack one or more biosynthetic *nifE* and *nifN* genes that are necessary for assembly of the FeMo-cofactor, suggesting that early Mo-dependent nitrogenases may have utilized an alternate pathway for Mo-usage predating the FeMo-cofactor. Our results underscore the profound impacts that protein-level innovations likely had on shaping global biogeochemical cycles throughout Precambrian, in contrast to organism-level innovations which characterize Phanerozoic eon.

KEYWORDS

ancestral sequence reconstruction, cofactor pocket, metalloenzyme, nitrogenase, nitrogen fixation

1 INTRODUCTION

All known life requires nitrogen for the synthesis of essential biomolecules, including nucleotides and amino acids. Though the atmosphere contains nearly 80% N_2 by volume, most organisms are not able to assimilate N_2 due to the enormous energetic cost of breaking the N≡N bond (MacKay & Fryzuk, 2004). Select bacteria and archaea called diazotrophs accomplish biological nitrogen fixation by nitrogenase metalloenzymes (E.C. 1.18.6.1), which catalyze the reduction of N_2 to bioavailable NH_3. Nitrogenases are an ancient family of enzymes; oldest isotopic biosignatures interpreted as evidence of nitrogenase activity date to ~3.2 billion years (Ga) (Stueken, Buick, Guy, & Koehler, 2015). Because nitrogen has been suggested to be an important limiting nutrient on geologic timescales (Falkowski, 1997), nitrogenases have likely played a key role in the expansion of the biosphere since the Archean.

The nitrogenase family consists of three homologous isozymes (Boyd, Hamilton, & Peters, 2011; Raymond, Siefert, Staples, & Blankenship, 2004) named for the differential metal content of the active-site cofactor: Mo-nitrogenase (Nif, encoded by *nif*), V-nitrogenase (Vnf, encoded by *vnf*), and Fe-nitrogenase (Anf, encoded by *anf*) (Bulen & LeComte, 1966; Eady, 1996; Joerger & Bishop, 1988; Mus, Alleman, Pence, Seefeldt, & Peters, 2018) (**Figure 1**). Of the three isozymes, Mo-nitrogenases are the most common and widely studied; V- and Fe-nitrogenases are comparatively rarer and only known in taxa that also possess Mo-nitrogenase (Boyd, Hamilton, et al., 2011; Dos Santos, Fang, Mason, Setubal, & Dixon, 2012). All three nitrogenase isozymes are structurally and functionally similar, each containing two protein components: the electron delivery component (NifH, VnfH, or AnfH) is a homodimer and the catalytic component is either an $\alpha_2\beta_2$ heterotetramer (MoFe protein, NifDK) or an $\alpha_2\beta_2\gamma_2$ heterohexamer (VFe protein, VnfDGK

or FeFe protein, AnfDGK) (**Figure 1a**) (Bulen & LeComte, 1966; Hales, Case, Morningstar, Dzeda, & Mauterer, 1986; Schmid et al., 2002; D. Sippel & Einsle, 2017). During catalysis, the electron delivery component transiently associates with and delivers electrons to the catalytic component (Hageman & Burris, 1978). Electrons accumulate at the active site for N_2 binding and reduction (Hoffman, Lukoyanov, Yang, Dean, & Seefeldt, 2014), which houses a homocitrate-metallocluster cofactor unique to each isozyme: the FeMo-cofactor in Nif (**Figure 1b**), FeV-cofactor in Vnf (**Figure 1c**), and FeFe-cofactor in Anf (**Figure 1d**) (Eady, 1996; Harris, Lukoyanov, et al., 2018; Krahn et al., 2002; Mus et al., 2018; D. Sippel & Einsle, 2017; Spatzal et al., 2011) Available spectral evidence suggest that these cofactors are structurally similar except for the substitution of a Mo, V, or additional Fe atom (Eady, 1996; Krahn et al., 2002; Daniel Sippel et al., 2018; Spatzal et al., 2011). Nevertheless, biochemical studies demonstrate variable catalytic properties among the three nitrogenase isozymes, including differential abilities to reduce alternative substrates (Harris, Lukoyanov, et al., 2018; Harris, Yang, Dean, Seefeldt, & Hoffman, 2018; B. Hu et al., 2018; Y. Hu, Lee, & Ribbe, 2011; Zheng et al., 2018). These catalytic variations likely arise due to a combination of the aforementioned cofactor compositional differences as well as differences in the surrounding protein environment (Fixen et al., 2016; Harris, Yang, et al., 2018; Lee et al., 2018; Rebelein, Lee, Newcomb, Hu, & Ribbe, 2018; Zheng et al., 2018).

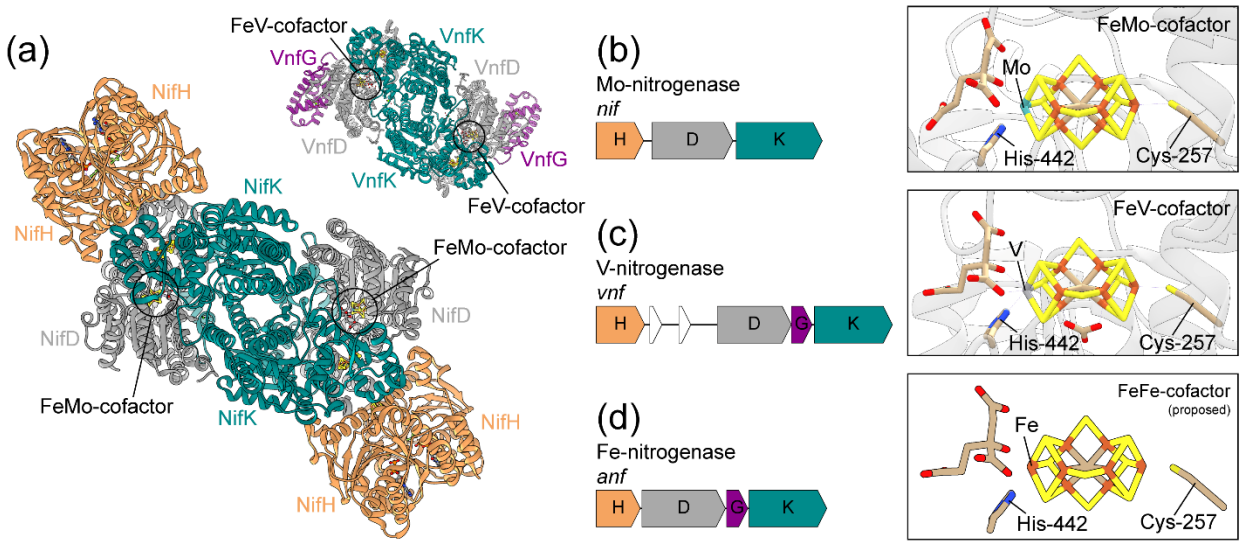


Figure 1. Structure and genetics of the three nitrogenase isozymes. (a) Structure of the *A. vinelandii* Mo-nitrogenase enzyme complex (NifHDK; PDB 1M34 (Schmid et al., 2002)) and V-nitrogenase VFe-protein component (VnfDGK; PDB 5N6Y (D. Sippel & Einsle, 2017)) (Fe-nitrogenase structure not previously published). The active-site FeMo-cofactor of Mo-nitrogenase and FeV-cofactor of V-nitrogenase are circled. (b-d) Catalytic genes and cofactor structures of *A. vinelandii* Mo-nitrogenase (b; PDB 3U7Q (Spatzal et al., 2011)), V-nitrogenase (c; PDB 5N6Y (D. Sippel & Einsle, 2017)), and Fe-nitrogenase (d; proposed structure (Harris, Lukoyanov, et al., 2018)). Cofactor atom coloring is as follows: C, tan; Fe, rust; Mo, cyan; N, blue; O, red; S, yellow.

Metal cofactor incorporation in nitrogenases is constrained at multiple levels. At the level of single enzyme functionality, nitrogenase biochemical and biophysical properties shape metal binding specificity. At a higher level, constraints arise from the partner proteins that constitute the biosynthetic mechanism for active-site cofactor assembly and insertion, best studied in Mo-nitrogenases (Curatti et al., 2007; Y. Hu & Ribbe, 2011; Rubio & Ludden, 2008). In the *Azotobacter vinelandii* (*A. vinelandii*) model system, FeMo-cofactor assembly requires several

95 associated proteins encoded within the *nif* gene cluster (Curatti et al., 2007; Dos Santos et al.,
96 2012; Shah, Allen, Spangler, & Ludden, 1994; Shah, Imperial, Ugalde, Ludden, & Brill, 1986; St
97 John et al., 1975; Tal, Chun, Gavini, & Burgess, 1991). Most notably, the biosynthetic *nifB, nifE,*
98 and *nifN* genes have been demonstrated, in addition to the catalytic *nifHDK* genes, to be — perhaps
99 minimally — required for FeMo-cofactor assembly and Mo-nitrogenase function (Curatti et al.,
100 2007; Dos Santos et al., 2012; Shah et al., 1994; Shah et al., 1986; St John et al., 1975; Tal et al.,
101 1991). In *A. vinelandii, nifE* and *nifN* loci are located just downstream of the *nifHDK* cluster,
102 whereas the *nifB* locus is located within a separate *nif* region near other regulatory and biosynthetic
103 *nif* genes (Setubal et al., 2009). NifB catalyzes the formation of a Fe-S-C metallocluster, a
104 precursor that forms the core of the mature FeMo-cofactor (Allen, Chatterjee, Ludden, & Shah,
105 1995; Y. Hu & Ribbe, 2011; St John et al., 1975). This precursor metallocluster is then transferred
106 to a protein heterotetramer composed of NifE and NifN subunits (Allen et al., 1995; Roll, Shah,
107 Dean, & Roberts, 1995), homologous to NifD and NifK, respectively, and likely having arisen by
108 gene duplication (Boyd, Anbar, et al., 2011). Within NifEN, the precursor is further modified via
109 the addition of homocitrate and Mo, and the mature cofactor is subsequently transferred to the
110 nitrogenase NifDK catalytic protein component (Roll et al., 1995). Unlike that for the FeMo-
111 cofactor, the biosynthetic pathways for the formation of the FeV- and FeFe-cofactors are relatively
112 unknown. However, transcriptional profiling of the three nitrogenase systems in *A. vinelandii*
113 suggests that FeV- and FeFe-cofactor synthesis relies on several *nif* genes in addition to *vnf* and
114 *anf* genes, respectively (Hamilton et al., 2011; Joerger & Bishop, 1988; Kennedy & Dean, 1992).
115 These include *nifBEN*, with the exception of certain taxa (including *A. vinelandii*) that possess
116 *vnfEN* homologs of *nifEN* that likely perform a similar biosynthetic function (Boyd, Anbar, et al.,
117 2011; Hamilton et al., 2011).

Paleobiological interest in nitrogenases has primarily centered on the coevolution of nitrogenase metal usage and the geochemical environment, with the possible implication that marine metal availabilities have significantly constrained nitrogenase evolution over geologic time (Anbar & Knoll, 2002; Boyd, Hamilton, et al., 2011; Canfield, Glazer, & Falkowski, 2010; Raymond et al., 2004). Inferences of ancient nitrogenase metal usage have relied on isotopic biosignatures (Stueken et al., 2015) and metal abundances (Anbar & Knoll, 2002) evidenced by the geologic record, as well as on phylogenetic reconstructions of both catalytic and cofactor biosynthesis proteins (Boyd, Anbar, et al., 2011; Boyd, Hamilton, et al., 2011; Raymond et al., 2004). High marine Fe concentrations and potential Mo scarcity prior to increased atmospheric oxygenation surrounding the ~2.3-2.5 Ga Great Oxidation Event (Anbar et al., 2007; Lyons, Reinhard, & Planavsky, 2014) has led to the hypothesis that Fe- or V- nitrogenases may have been dominant in early oceans (Anbar & Knoll, 2002; Canfield et al., 2010) and possibly predate Mo-nitrogenases (Raymond et al., 2004). More recent phylogenetic reconstructions have instead suggested that the evolution of Mo-nitrogenases, dated by time-calibrated phylogenies of Nif/Vnf/AnfDKEN sequences to ~1.5–2.2 Ga (Boyd, Anbar, et al., 2011), preceded that of V- and Fe-nitrogenases (Boyd, Hamilton, et al., 2011). These phylogenetic inferences are also consistent with the observation that *vnf* and *anf* genes are only present in organisms that also harbor *nif*, and that V-/Fe-nitrogenase assembly relies on *nif* biosynthetic genes (Hamilton et al., 2011; Joerger & Bishop, 1988; Kennedy & Dean, 1992). However, ~3.2-Ga isotopic signatures of biological nitrogen fixation suggest an earlier origin of nitrogenase (Stueken et al., 2015), and, even though isotopically consistent with Mo-dependent N-fixation, predate age estimates of both Mo-nitrogenase (Boyd, Anbar, et al., 2011) and earliest marine Mo availability (Anbar et al., 2007;

141 Anbar & Knoll, 2002; Lyons et al., 2014). Thus, the evolutionary trajectory of nitrogenase metal
142 usage — and by extension the link between nitrogenase evolution and marine metal availabilities
143 over geologic time — is not yet known.
144
145 Here, we explored the indicators of nitrogenase metal usage history by a combinatorial method
146 relying on ancestral sequence reconstruction, a method by which inferred, historical protein
147 sequence information can be linked to functional molecular properties evidenced by computed
148 structures or laboratory experiments (Aadland, Pugh, & Kolaczkowski, 2019; Benner, Sassi, &
149 Gaucher, 2007; Thornton, 2004). These paleogenetic approaches have been increasingly applied
150 in biogeochemically relevant molecular studies to offer insights into the coevolution of life and
151 Earth (Garcia & Kacar, 2019; Gomez-Fernandez et al., 2018; Kacar, Hanson-Smith, Adam, &
152 Boekelheide, 2017; Shih et al., 2016). We reconstructed the phylogenetic history of Mo-, V-, and
153 Fe-nitrogenases in order to resurrect ancestral nitrogenases *in silico*, as well as to map the
154 taxonomic distribution of cofactor biosynthetic components considered necessary for Mo-
155 dependence (Curatti et al., 2007; Shah et al., 1994; Shah et al., 1986; St John et al., 1975; Tal et
156 al., 1991). By this combined approach, we find phylogenetic and ancestral sequence features
157 suggestive of Mo-dependence, potentially by an alternate pathway predating the origin of the
158 FeMo-cofactor. We speculate that this unknown and possibly transient pathway may today be
159 present in basal nitrogenase lineages. Integration of protein evolution and paleobiology is a unique
160 melding of disparate data sets and may allow construct-and-build hypotheses that address
161 interactions ranging from the external environment to the cellular environment, and from the
162 cellular environment to the that maintained around the interacting protein. The exchange of
163 materials across these different scales necessitates constraints on the flow and availability of

substrates that make such exchanges possible. It is the specific nature of these constraints and how they may change in response to external perturbations that enable us to develop completely new testable hypothesis that connect geochemical reservoirs with biological metabolisms — hypotheses that cannot be constructed from macroevolutionary or geological frameworks alone.

2 MATERIALS AND METHODS

2.1 Ancestral reconstruction of nitrogenase protein sequences

An initial dataset of extant nitrogenase Nif /Vnf/AnfHDK homologs was constructed by retrieving amino acid sequences from the National Center for Biotechnology Information non-redundant protein database, accessed September 2018 (O'Leary et al., 2016). Potential homologs were identified by BLASTp (Camacho et al., 2009) using query sequences from *A. vinelandii* (NifH: Avin_01380, NifD: Avin_01390, NifK: Avin_01400) and an Expect value cutoff of <1e-5. The dataset was then manually curated to remove partial and distantly related sequences. Additional nitrogenase sequences were manually retrieved from the Joint Genome Institute Integrated Microbial Genomes and Microbiomes database, accessed September 2018 (Chen et al., 2019). The nitrogenase sequence dataset was finalized to include NifHDK sequences from 256 taxa, AnfHDK sequences from 14 taxa, VnfHDK sequences from 14 taxa, and outgroup light-independent protochlorophyllide oxidoreductase (Bch/ChlLNB) sequences — sharing distant homology with nitrogenases (Boyd, Anbar, et al., 2011; Y. Hu & Ribbe, 2015; Raymond et al., 2004) — from 10 taxa (Appendix S1; additional analyses were performed with an expanded outgroup, Appendix S2). Only one Nif/Anf/VnfHDK sequence set was retained per genus to broaden taxonomic sampling. Equal sequence sampling for Anf and Vnf was made to remove the potential for oversampling bias in ancestral sequence inference. H-, D-, and K-subunit sequences corresponding

187 to each taxon were manually checked for synteny of their encoding genes. AnfHDK and VnfHDK
188 sequences were identified by the proximity of each gene locus to *anfG* or *vnfG*, which encodes the
189 additional G-subunit present in the VeFe or FeFe protein, respectively, but not present in the MoFe
190 protein (Eady, 1996). Finally, the presence of the cofactor biosynthetic *nifBEN* genes was
191 investigated for all taxa represented in our dataset by BLASTp, as well as by manually inspecting
192 the *nif* genome region.
193
194 Reconstruction of ancestral nitrogenase sequences was performed by PhyloBot (Hanson-Smith &
195 Johnson, 2016) (www.phylobot.com), which automates multiple sequence alignment,
196 phylogenetic reconstruction, and ancestral sequence inference methods. The concatenated 294-
197 sequence dataset of Nif/Anf/VnfHDK homologs (including 10 Bch/ChlLNB outgroup sequences)
198 was aligned by MSAProbs v0.9 5r1 (Liu, Schmidt, & Maskell, 2010) and MUSCLE v3.8.31
199 (Edgar, 2004). Both alignment outputs were then used to perform phylogenetic reconstruction by
200 RAxML v8.1.15 (Stamatakis, 2014) under 6 different combinations of amino acid substitution and
201 rate heterogeneity models. Branch support was evaluated by the approximate likelihood ratio test
202 (aLRT) (Anisimova & Gascuel, 2006), which assesses the gain in overall likelihood against a null
203 hypothesis of branch length = 0. Additional phylogenetic reconstructions with an expanded
204 outgroup were performed outside of Phylobot to resolve root positioning, but were not used in
205 subsequent ancestral sequence inference (Appendix S2). Ancestral sequences were inferred by
206 joint maximum likelihood using CODEML v4.2 (Z. Yang, 2007) at all nodes within the 12
207 Phylobot-constructed phylogenies (Tree-1 – Tree-12), with gaps inferred by parsimony. To assess
208 ancestral sequence robustness to phylogenetic uncertainty (Hanson-Smith, Kolaczkowski, &
209 Thornton, 2010), ancestors inferred from the top five phylogenies ranked by log likelihood scores

210 were selected for further analysis (Table 1, Appendix S2). Finally, to evaluate the effects of
211 ambiguously reconstructed sites on subsequent structural analyses, Bayesian sampled ancestors
212 were inferred from the maximum likelihood site posterior probabilities calculated by CODEML
213 (Aadland et al., 2019). 100 random Bayesian sequence were generated for each of five ancestral
214 nodes of interest across the top five phylogenies. Thus, 25 maximum likelihood and 2,500
215 Bayesian-sampled ancestral sequences were analyzed in total. All maximum likelihood
216 reconstructed trees and ancestral sequences are available for view and download at
217 http://phylobot.com/613282215/.

Table 1. Alignment and evolutionary model parameter combinations for the top five phylogenies, ranked by log likelihood scores.

Phylogeny	Alignment method	Evolutionary model	Log likelihood	Maximum likelihood ancestors
Tree-1	MSAProbs	CAT + LG	-300069.08	AncA-1 – AncE-1
Tree-2	MSAProbs	CAT + WAG	-303296.08	AncA-2 – AncE-2
Tree-3	MSAProbs	Γ + LG	-303951.21	AncA-3 – AncE-3
Tree-4	MUSCLE	CAT + LG	-304457.52	AncA-4 – AncE-4
Tree-5	MSAProbs	Γ + WAG	-305229.49	AncA-5 – AncE-5

References: (Le & Gascuel, 2008; Quang, Gascuel, & Lartillot, 2008; Whelan & Goldman, 2001; Z. Yang, 1993)

227 **2.2 Structural homology modeling of extant and ancestral nitrogenase D-subunits**

228 Structural homology modeling of 33 extant and ancestral (25 maximum likelihood and 2,500
229 Bayesian-sampled) nitrogenase D-subunit proteins was performed by Modeller v9.2 (Sali &
230 Blundell, 1993). Extant nitrogenase sequences, broadly sampled from the reconstructed
231 nitrogenase phylogeny, were modeled to provide comparisons with ancestral models. D-subunit
232 sequences from extant and ancestral nitrogenases were aligned to 38 NifD and 2 VnfD structural
233 templates retrieved from the Protein Data Bank (Berman et al., 2000), accessed November 2018
234 (Appendix S5; published AnfD models not available at time of analysis). Information from all 40
235 templates was used to model each structure. All models were generated by specifying the inclusion
236 of the FeMo-cofactor of the 3U7Q NifD structure (Spatzal et al., 2011), selected as the highest
237 resolution Mo-nitrogenase template. To assess the effect of the template cofactor type on the
238 generated structure, additional models were constructed by specifying the inclusion of the FeV-
239 cofactor of the 56NY VnfD template (D. Sippel & Einsle, 2017) (Appendix S5). 100 modeling
240 replicates were performed per sequence and assessed by averaging over the scaled Modeller
241 objective function, Discrete Optimized Protein Energy, and high resolution Discrete Optimized
242 Protein Energy scores, as previously described (Aadland et al., 2019). The ten best modeling
243 replicates per extant sequence, ten best replicates per maximum likelihood ancestral sequence, and
244 the single best replicate per Bayesian-sampled variant sequence were selected for further analysis,
245 totaling 3,080 models with the FeMo-cofactor specified.

246

247 **2.3 Active-site pocket volume calculation of extant and ancestral D-subunit models**

248 Volumes of the modeled ancestral and extant D-subunit active-site cofactor pockets were
249 calculated by POVME v2.0 (Durrant, Votapka, Sorensen, & Amaro, 2014). Spatial coordinates

250 and the inclusion region for volume calculation were specified manually. Pocket volumes were
251 calculated with a grid spacing of 0.5 Å and a 1.09 Å distance cutoff from any receptor atom's van
252 der Waals radius. Volume outside of the modeled convex hull of the cofactor pocket as well as
253 noncontiguous volume were removed. Statistical analysis of ancestral and extant pocket volume
254 data was performed in R (R Core Team, 2014).
255

256 **3 RESULTS**

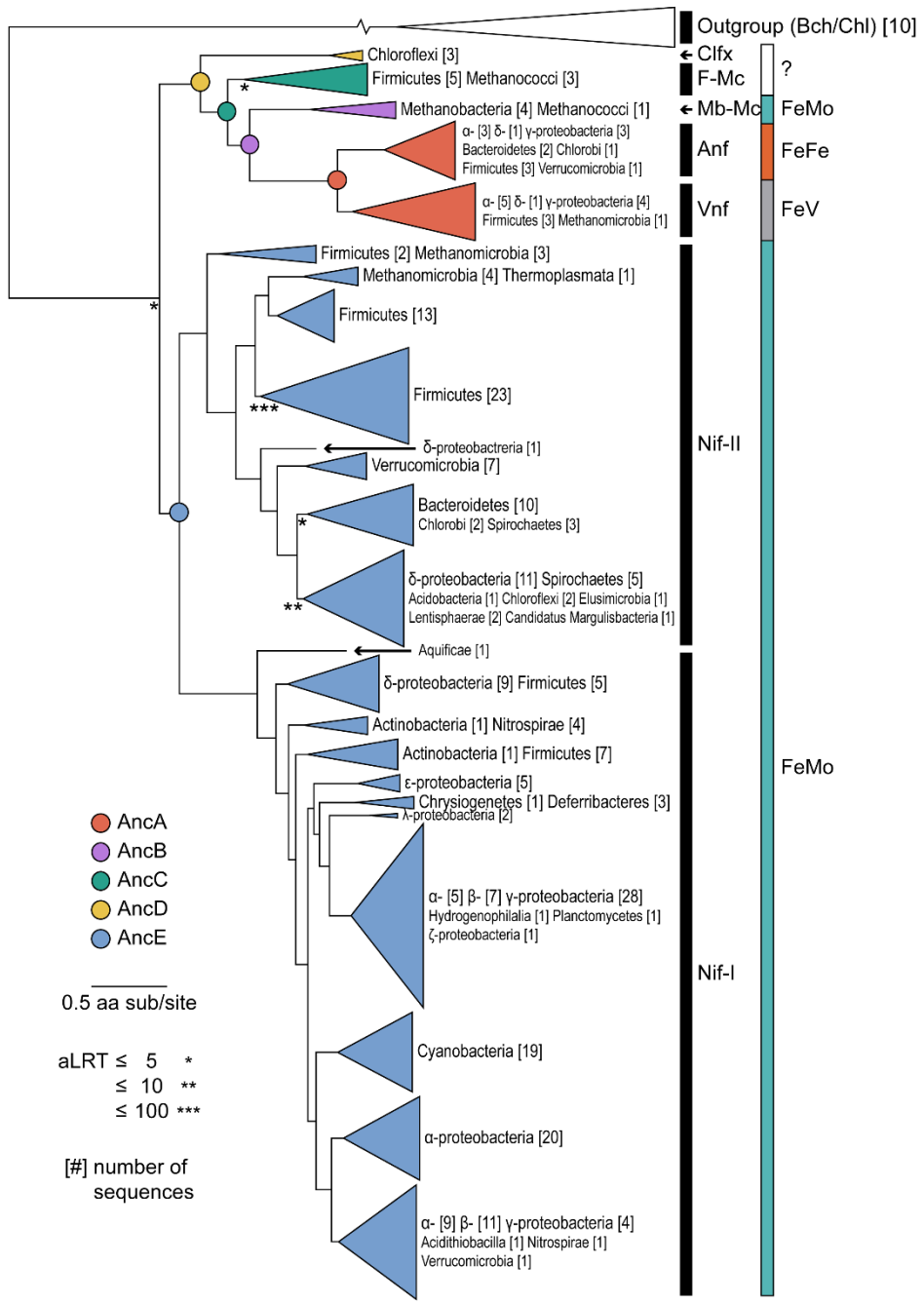
257 **3.1 V- and Fe-nitrogenases diversified after Mo-nitrogenases**

258 We reconstructed the phylogenetic history of Mo-, V-, and Fe-nitrogenases and to infer ancestral
259 nitrogenase sequences and associated indicators of nitrogenase metal dependence.
260 Nif/Anf/VnfHDK protein homologs curated from the National Center for Biotechnology
261 Information and Joint Genome Institute databases represent 20 bacterial and archaeal phyla, 11 of
262 which are known from experimental investigations to include diazotrophic taxa (Dos Santos et al.,
263 2012; Ormeño-Orrillo, Hungria, & Martinez-Romero, 2013) (Appendix S1). The five most
264 represented phyla in our dataset — Bacteroidetes, Cyanobacteria, Firmicutes, Proteobacteria, and
265 Euryarchaeota — encompass ~80% of the curated sequences. Our dataset also presents genomic
266 evidence of nitrogen fixation within the Acidobacteria, Actinobacteria, Aquificae, Chlorobi,
267 Chloroflexi, Chrysiogenetes, Deferribacteres, Elusimicrobia, Fusobacteria, Lentisphaerae,
268 *Candidatus* Margulisbacteria, Nitrospirae, Planctomycetes, Spirochaetes, and Verrucomicrobia.
269

270 Our maximum likelihood nitrogenase phylogeny segregates Nif/Anf/VnfHDK sequences into two
271 major lineages similarly observed in previous studies (Boyd, Anbar, et al., 2011; Raymond et al.,
272 2004) (Tree-1; **Figure 2**): the first comprises Nif-I and Nif-II Mo-nitrogenases (aerobic/facultative

273 bacteria and anaerobic/facultative bacteria/archaea, respectively; highlighted in blue) (Raymond
274 et al., 2004), and the second comprises V- and Fe-only-nitrogenases (Vnf and Anf, respectively;
275 highlighted in red) as well as three clades of "uncharacterized" nitrogenases that lack extensive
276 experimental characterization with regard to metal dependence (Mb-Mc, highlighted in purple; F-
277 Mc, highlighted in green; Clfx, highlighted in yellow) (Boyd, Hamilton, et al., 2011; Dekas,
278 Poretsky, & Orphan, 2009; Dos Santos et al., 2012; Mehta & Baross, 2006). We additionally
279 analyzed four alternate phylogenies (ranked by log likelihood scores), reconstructed by varying
280 alignment methods, amino acid substitution models, and rate heterogeneity models (Tree-2–Tree-
281 5; Appendix S2). The aforementioned major nitrogenase clades are present across all alternate
282 phylogenetic topologies. Branch ordering of these major clades is similarly consistent, with the
283 exception of root position: in Tree-3 and Tree-5, the root is placed between Nif-I and Nif-
284 II/uncharacterized/Anf/Vnf, rather than between Nif-I/Nif-II and uncharacterized/Anf/Vnf as in
285 Tree-1, Tree-2, and Tree-4. Nevertheless, additional phylogenetic analyses incorporating an
286 expanded outgroup provide stronger support for the root position of Tree-1, Tree-2, and Tree-4
287 (Appendix S2).
288
289 In all analyzed nitrogenase phylogenies, Vnf and Anf sequences (highlighted in red) form
290 reciprocally monophyletic clades that branch immediately distal to Mb-Mc nitrogenases
291 (highlighted in purple) (**Figure 2**, Appendix S2). The reciprocal monophyly of Vnf and Anf, as
292 well as the clustering of Mb-Mc, Vnf, and Anf, is well-supported across all phylogenetic
293 topologies (aLRT > 10^6). The Mb-Mc clade is composed of archaeal hydrogenotrophic
294 methanogens within classes Methanobacteria and Methanococci. Because all Mb-Mc taxa possess
295 *nifBEN* genes demonstrated to be necessary for the synthesis of the FeMo-cofactor (Curatti et al.,

296 2007; Dos Santos et al., 2012; Shah et al., 1994; Shah et al., 1986; St John et al., 1975; Tal et al.,
297 1991), it is likely that Mb-Mc nitrogenases are Mo-dependent (Boyd, Hamilton, et al., 2011). Thus,
298 the phylogenetic positioning of Vnf and Anf is consistent with previous suggestions that V- and
299 Fe-nitrogenases diversified after Mo-nitrogenases (Boyd, Hamilton, et al., 2011).
300



301

302 **Figure 2.** Maximum likelihood phylogeny of concatenated Nif/Anf/VnfHDK nitrogenase and
303 Bch/ChlLNB outgroup protein sequences (Tree-1; see Table 1). Ancestral nodes analyzed in this
304 study are labeled AncA–AncE. Known active-site cofactor metal content is listed on the right.
305 Branch support is derived from the approximate likelihood ratio test (aLRT). Branch length scale
306 is in units of amino acid substitutions per site. Outgroup branch break used to conserve space; true
307 branch length = 5.578 substitutions per site. Phylogeny coloring is as follows: Clfx, yellow; F-Mc,
308 green; Mb-Mc, purple; Anf/Vnf, red; Nif-I/-II, blue.
309

310 **3.2 Basal uncharacterized nitrogenases lack associated genes necessary for FeMo-cofactor**
311 **synthesis**

312 In addition to investigating the phylogenetic relationships between Mo-, V-, and Fe-nitrogenase
313 isozymes, we mapped the presence of the biosynthetic *nifB*, *nifE*, and *nifN* genes — necessary for
314 FeMo-cofactor assembly (Curatti et al., 2007; Dos Santos et al., 2012; Shah et al., 1994; Shah et
315 al., 1986; St John et al., 1975; Tal et al., 1991) — among taxa represented in our dataset. All
316 analyzed taxa possess the full complement of *nifBEN* biosynthetic genes, with the exception of
317 two uncharacterized clades: Clfx (highlighted in yellow) and F-Mc (highlighted in green) (**Figure**
318 **2**). Within the lineage containing Vnf and Anf nitrogenases, Clfx and F-Mc clades are most basal.
319 These branching positions of Clfx and F-Mc clades within the Vnf/Anf lineage are consistently
320 observed across all phylogenetic topologies (Trees-1–5; Appendix S2) and are well-supported
321 (aLRT > 10^3 for Clfx, aLRT > 10^{18} for F-Mc). In Tree-1-, 2-, and -4, as well as in trees
322 reconstructed with an expanded outgroup (Appendix S2), Clfx and F-Mc clades also branch
323 immediately distal to the root.
324

325 Both basal Clfx and F-Mc nitrogenases are mostly derived from thermophilic hosts, and lack one
326 or more associated *nifEN* cofactor biosynthesis genes. Clfx nitrogenases, constituting the most
327 basal uncharacterized clade within the uncharacterized/Vnf/Anf lineage, are in the present dataset
328 derived from three mesophilic or thermophilic Chloroflexi species (**Figure 2**, highlighted in
329 yellow). The *nif* clusters of Clfx taxa lack both *nifE* and *nifN* genes, and the typically continuous
330 *nifHDK* genes observed in other taxa are instead interrupted by *nifB* (arranged *nifHBDK*)(Setubal
331 et al., 2009). F-Mc nitrogenases branch immediately distal to the Clfx clade and represent eight
332 thermophilic Firmicutes and archaeal methanogen (class Methanococci) taxa (**Figure 2**,
333 highlighted in green). F-Mc species possess biosynthetic *nifB* and *nifE* genes, but not *nifN*, with
334 the exception of *Methanothermococcus thermolithotrophicus* that retains *nifN*. A previous study
335 found that sequence features and modeled structural features of F-Mc nitrogenases closely
336 resemble those of Mo-nitrogenases (McGlynn, Boyd, Peters, & Orphan, 2012). However, the
337 absence of *nifN* in most F-Mc taxa suggests that these strains are not capable of synthesizing the
338 FeMo-cofactor. Though the lack of *nifE* and/or *nifN* genes in Clfx and F-Mc taxa might
339 additionally suggest that such strains cannot express functional nitrogenases, some have been
340 experimentally observed to fix nitrogen: the Clfx species *Oscillochloris trichoides* (lacking *nifEN*)
341 (Keppen, Lebedeva, Troshina, & Rodionov, 1989; Kuznetsov et al., 2011) and the F-Mc species
342 *Methanocaldococcus sp.* FS406-22 (lacking *nifN*) (Keppen et al., 1989; Kuznetsov et al., 2011;
343 Mehta & Baross, 2006), in addition to uncharacterized anaerobic methane-oxidizing archaea not
344 included in the present study (Dekas et al., 2009). The ability of early-branching Clfx and F-Mc
345 nitrogenases to fix nitrogen in the absence of *nifEN* genes may indicate an atypical pathway for
346 cofactor assembly and incorporation not used for extant Mo-, V-, and Fe-nitrogenases.
347

348 **3.3 High statistical support for ancestral nitrogenase active-site residues**

349 We inferred ancestral sequences for each of the H-, D-, and K-subunits that constitute the
350 nitrogenase enzyme complex (**Figure 1**) across five phylogenetic topologies (Tree-1–5; Appendix
351 S2). Ancestral nitrogenase sequences were inferred for five well-supported internal nodes along a
352 phylogenetic transect between Mo- (highlighted in blue) and V-/Fe-nitrogenases (highlighted in
353 red) (**Figure 2**). The five targeted nodes are: AncA (ancestral to Anf and Vnf), AncB (ancestral to
354 AncA and Mb-Mc), AncC (ancestral to AncB and F-Mc), AncD (ancestral to AncC and Clfx), and
355 AncE (ancestral to Nif-I and Nif-II). Thus, AncA–D are nested, whereas AncE lies along a
356 divergent lineage toward Nif-I and Nif-II Mo-nitrogenases. For further analyses, we selected the
357 maximum likelihood ancestral sequence per target node from each of the five phylogenies (Tree-
358 1–5), totaling 25 sequences. Due to differences in root position, identical AncE nodes were not
359 present across all topologies and analogous nodes were instead selected (Appendix S2). Ancestral
360 sequences are hereafter labeled with the tree likelihood rank from which they were inferred (e.g.,
361 AncA from Tree-1 is labeled AncA-1). All tree and ancestral sequence information can be found
362 at http://phylobot.com/613282215/.

363

364 Mean site posterior probabilities for ancestral nitrogenase HDK sequences across all phylogenies
365 range between ~0.83 and 0.91, and for the highest-likelihood phylogeny (Tree-1), between ~0.84
366 and 0.90 (Appendix S3). Ancestral sequence support generally decreases with increasing
367 phylogenetic node age. For example, within the uncharacterized/V-/Fe-nitrogenase linage, AncA-
368 1 has the highest mean posterior probability (0.90 ± 0.18) and AncD-1 has the lowest mean
369 posterior probability (0.84 ± 0.22). Mean ancestral sequence probability for each node also does
370 not deviate by more than ~0.02 across each of the five phylogenetic topologies (Tree-1–5;

371 Appendix S3). These observations suggest that sequence support for ancestral nitrogenases is more
372 sensitive to ancestral node position than to topological differences between the analyzed trees.
373
374 In addition to surveying total ancestral HDK sequence support, we analyzed support for 30 active-
375 site residues, defined as those residing within 5 Å of any atom in either the FeMo-cofactor of the
376 *A. vinelandii* NifD protein (PDB 3U7Q (Spatzal et al., 2011)) or the FeV-cofactor of the *A.*
377 *vinelandii* VnfD protein (PDB 5N6Y (D. Sippel & Einsle, 2017)) (**Figure 3**). These active-site
378 residues are not contiguous but are instead scattered throughout the D-subunit sequence. Mean
379 posterior probabilities of ancestral active-site residues, which range between 0.92 to 0.98 across
380 all phylogenies, are consistently greater than those of entire reconstructed nitrogenase HDK
381 sequences (0.83–0.91) (Appendix S3). Of the 30 active-site residues, only five sites have, in one
382 or more ancestral sequences, plausible alternative reconstructions with posterior probabilities >
383 0.30: sites 59, 69, 358, 360, 425, 441 (site numbering both here and hereafter based on *A. vinelandii*
384 NifD). No ancestral sequences have more than three active-site residues with such plausible
385 alternative reconstructions. Ten active-site residues are conserved across all analyzed extant
386 nitrogenases: Val-70, Gln-191, His-195, Cys-275, Arg-277, Ser-278, Gly-356, Phe-381, Gly-424,
387 His-442. These conserved residues are thus reconstructed in all ancestral nitrogenases
388 unambiguously (site posterior probability = 1.00). Statistical support for ancestral active-site
389 residues (greater than 0.92) underpins subsequent analyses of ancestral active-site properties that
390 may inform inferences of nitrogenase metal dependence.

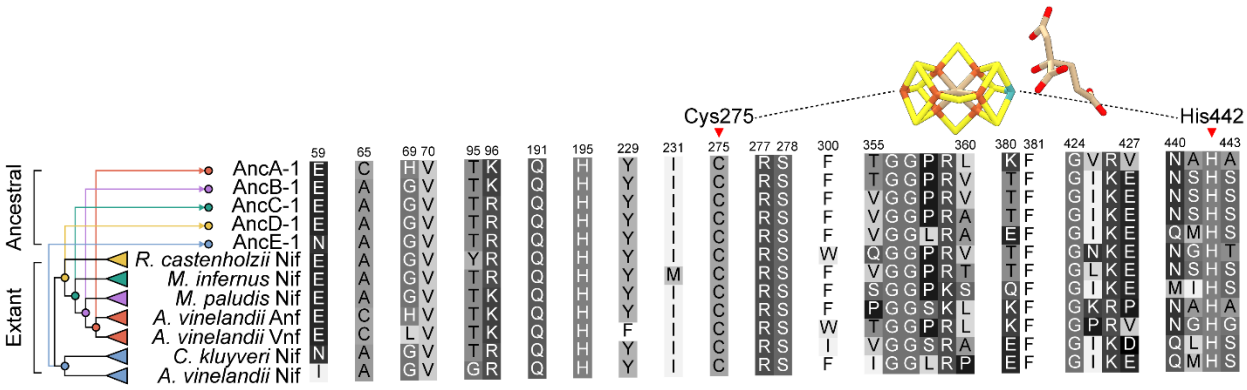
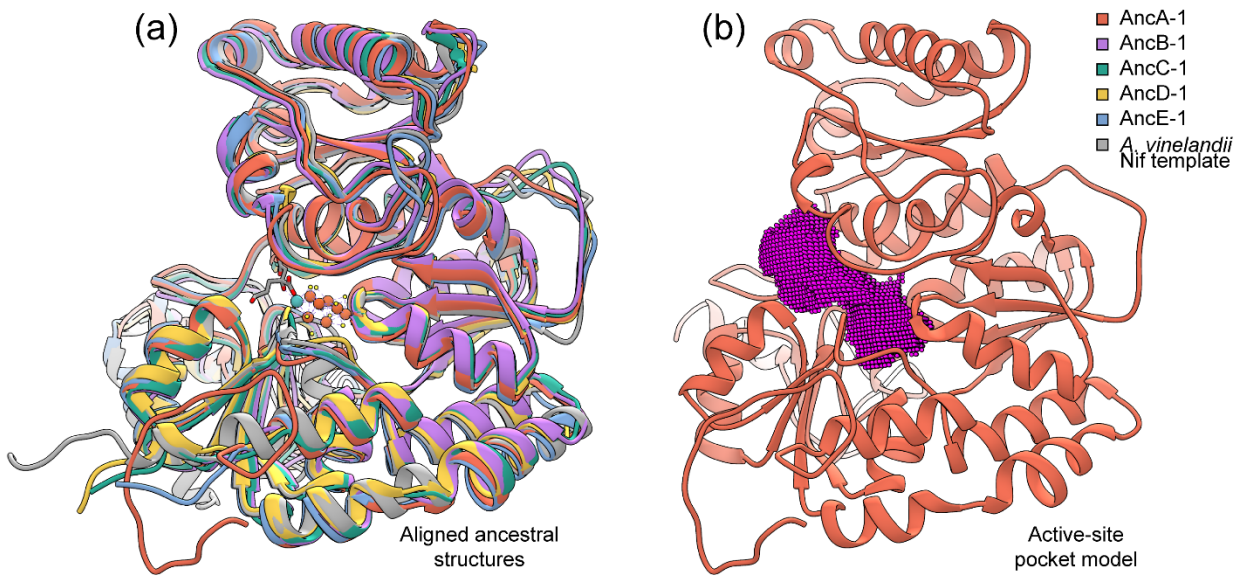


Figure 3. Active-site protein environment of representative ancestral and extant nitrogenases. All residues located within 5 Å of any atom in either the FeMo-cofactor of the *A. vinelandii* NifD protein (PDB 3U7Q (Spatzal et al., 2011)) or the FeV-cofactor of the *A. vinelandii* VnfD protein (5N6Y (D. Sippel & Einsle, 2017)). Residue numbering from aligned *A. vinelandii* NifD. Cys-275 and His-442 residues that coordinate the cofactor are indicated by red arrows. Phylogeny coloring is as follows: Clfx, yellow; F-Mc, green; Mb-Mc, purple; Anf/Vnf, red; Nif-I/-II, blue.

3.4 Active-site structural features are uninformative for inferring ancestral metal dependence

To investigate metal-specific features of ancestral nitrogenase structures, we generated homology models of both extant and ancestral nitrogenase D-subunits that house the active site (**Figure 4a**). First, we modeled 33 broadly sampled extant nitrogenase NifD, VnfD, and AnfD sequences to benchmark classifications of ancestral nitrogenase models. Second, we calculated structural models of 25 nitrogenase ancestors inferred by maximum likelihood and of 2,500 ancestors inferred by random Bayesian sampling of maximum likelihood site posterior probabilities (100 Bayesian samples per maximum likelihood ancestor). We generated ten model replicates per extant sequence and maximum likelihood sequence, and one model per Bayesian-sampled sequence. All structures were modeled with the FeMo-cofactor included (additional modeling runs were

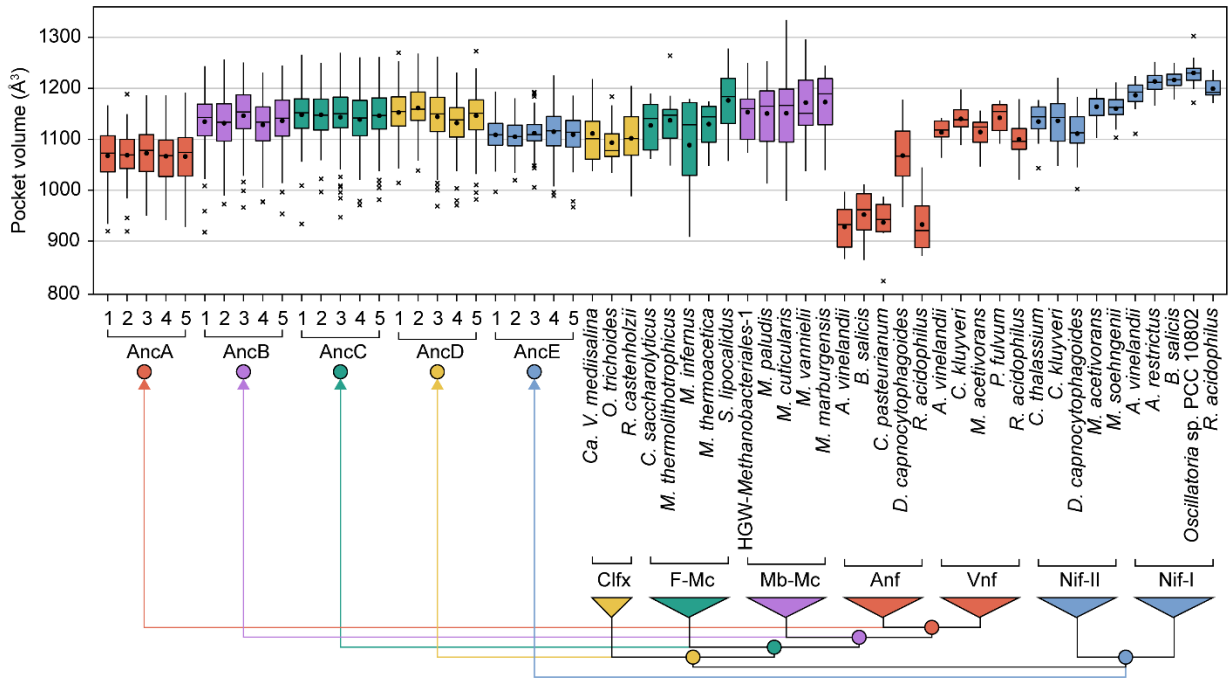
410 executed with the FeV-cofactor included; Appendix S5). In total, 3,080 models were generated
411 with the FeMo-cofactor.
412



413

414 **Figure 4**. Structural and active-site pocket modelling of ancestral nitrogenases (a) Modeled D-
415 subunit protein structures of ancestral nitrogenases inferred from the highest-likelihood phylogeny
416 (Tree-1; **Figure 2**) aligned to an *A. vinelandii* Nif structural template (PDB 3U7Q (Spatzal et al.,
417 2011)). (b) Example of a modeled active-site pocket for ancestral nitrogenase AncA-1. The 0.5-
418 Å-resolution point field generated for pocket volume calculation is shown in pink.
419

420 For each of the 3,080 extant and ancestral D-subunit nitrogenase models, we calculated the volume
421 of the active-site pocket (**Figure 4b**), a parameter previously used to classify the metal dependence
422 of extant uncharacterized nitrogenases (McGlynn et al., 2012). These pocket volume values are
423 plotted in **Figure 5**. Among modeled extant nitrogenases, mean pocket volumes are 1175.12 ±
424 51.93 $Å^3$ for Mo-nitrogenases, 1121.86 ± 36.36 $Å^3$ for V-nitrogenases, and 963.39 ± 75.80 $Å^3$ for
425 Fe-nitrogenases.



426

427 **Figure 5.** Extant and ancestral nitrogenase active-site pocket volumes. Pocket volumes calculated
428 for ancestral and representative extant nitrogenase D-subunit structures modeled with the FeMo-
429 cofactor. Each ancestral plot contains 110 volume calculations (ten model replicates per maximum
430 likelihood sequence plus one model for each of 100 Bayesian-sampled sequences) and each extant
431 plot contains 10 volume calculations (10 model replicates per extant sequence). Median values are
432 indicated by bars, mean values by points, the range (excluding outliers) by whiskers, and outliers
433 by crosses. Phylogeny coloring is as follows: Clfx, yellow; F-Mc, green; Mb-Mc, purple; Anf/Vnf,
434 red; Nif-I/-II, blue.

435

436 We observe less difference between mean pocket volumes of extant V-nitrogenases (1121.86 ±
437 36.36 Å^3) and Nif-II Mo-nitrogenases (1141.13 ± 46.30 Å^3) than between Nif-I (1209.11 ± 30.79
438 Å^3) and Nif-II Mo-nitrogenases. A statistical nonparametric test of volume median differences also
439 suggests greater similarity between V- and Nif-II Mo-nitrogenases than between Nif-I and Nif-II
440 nitrogenases (Appendix S5). All V-nitrogenase and Nif-II Mo-nitrogenase volume values range

441 between 1002.13 and 1220.63 Å^3, which problematically overlap with the volume ranges of AncB
442 (917.38–1256.25 Å^3), AncC (933.75–1269.50 Å^3), AncD (968.50–1272.75 Å^3), and AncE
443 (966.75–1225.25 Å^3) ancestral models, as well as ranges for uncharacterized Clfx (988.25–
444 1218.13 Å^3) and F-Mc (908.00–1277.75 Å^3) nitrogenases. The volume range of AncA models
445 (919.00-1191.13 Å^3) lies between the ranges of both V- (1020.75–1197.63 Å^3) and Fe-
446 nitrogenases (821.38–1177.00 Å^3). These same overall patterns are observed when comparing only
447 maximum likelihood ancestral models or Bayesian-sampled ancestral models, as well as alternate
448 modeling runs with the FeV-cofactor (Appendix S5). Greater similarities between V- and Nif-II
449 Mo-nitrogenases than between Nif-I and Nif-II Mo-nitrogenases suggests that, contrary to
450 previous analyses (McGlynn et al., 2012), modeled pocket volume may be uninformative for
451 inferring ancestral metal dependence. At the very least, the overlap in volume ranges between
452 ancestors and extant isozymes of varying metal dependence in our analyses preclude the
453 unambiguous classification of ancestral metal dependence by these structural features.
454

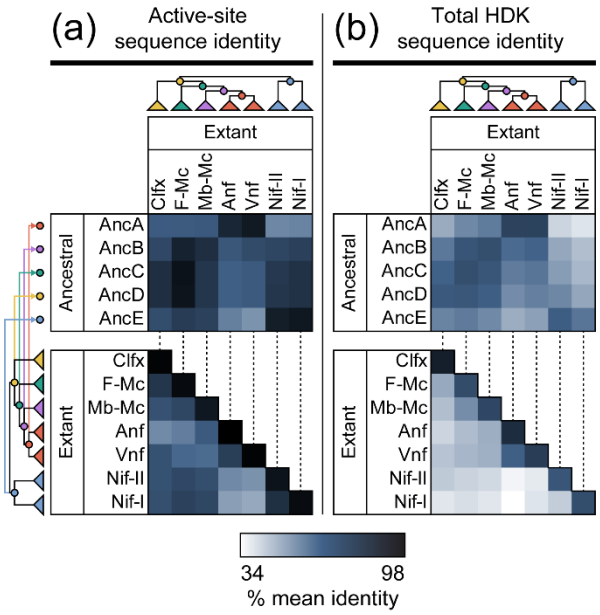
455 **3.5 Oldest ancestral nitrogenase active-site sequences resemble those of Mo-nitrogenases**

456 We analyzed sequence features of both ancestral and extant nitrogenases to identify those
457 correlated with metal dependence. In particular, we focused on nitrogenase active-site sequences
458 for three reasons: (1) active-site residues are known to affect catalytic efficiency and substrate
459 specificity (Brigle et al., 1987; Christiansen, Cash, Seefeldt, & Dean, 2000; Fixen et al., 2016;
460 Kim, Newton, & Dean, 1995; Sarma et al., 2010; Z. Y. Yang, Moure, Dean, & Seefeldt, 2012) and
461 thus may be tuned to nitrogenase metal dependence (2) active-site sequence features have
462 previously been used to classify the metal dependence of extant uncharacterized nitrogenases

463 (McGlynn et al., 2012) (3) ancestral active-site residues are reconstructed here with high statistical
464 support as compared to the total HDK sequence (see Section 3.3).
465
466 We first assessed the sensitivity of ancestral sequence variation to phylogenetic uncertainty and
467 ancestral statistical support. Overall, mean identities for ancestral sequences compared across
468 different nodes range from ~55 to 90%. Ancestral sequences inferred from the same node across
469 alternate phylogenies (Tree-1–5**)** have relatively high mean identities, ranging from ~93 to 95%
470 across the total HDK sequence and from ~96 to 100% within the active site. These high mean
471 identities suggest that topological differences among the alternate phylogenies used for ancestral
472 sequence inference do not contribute to a high degree of ancestral sequence variation. Identities
473 among sequences inferred from the same node also do not appear to be correlated with statistical
474 support. For example, though full AncA HDK sequences are reconstructed with the highest mean
475 statistical support (~0.89–0.91), they exhibit neither the lowest nor the highest mean identities as
476 compared with sequences inferred from other nodes (Appendix S3).
477
478 We next identified specific residues within the nitrogenase active site that are unique to particular
479 isozymes of known metal dependence in order to survey their occurrence in ancestral sequences.
480 We found that three active-site residues are unique to Mo-nitrogenases, six are unique to V-
481 nitrogenases, five are unique to Fe-nitrogenases, and six are unique to V- and Fe-nitrogenases
482 (Appendix S4). Surprisingly, most nitrogenase ancestors exhibit comparable numbers of residues
483 unique to either V-/Fe-nitrogenases or Mo-nitrogenases, and thus their occurrence does not appear
484 informative for inferring ancestral metal dependence (e.g., ancestral AncC sequences contain two
485 residues unique to V-/Fe-nitrogenases and two residues unique to Mo-nitrogenases). An exception

is AncA sequences, which contain a preponderance of active-site residues that are unique to extant V- and Fe-nitrogenases. One of the residues unique to V-nitrogenases, Thr-355, has recently been suggested to interact directly with a proposed carbonate ligand only present in the FeV-cofactor (D. Sippel & Einsle, 2017). This carbonate ligand lies within a protein loop that also contains Pro-358, unique to V-nitrogenases, and Leu-360, unique to V- and Fe-nitrogenases. These Thr-355, Pro-358, and Leu-360 residues are observed across all AncA sequences, reflecting specific sequence resemblances that may be associated with V-dependence.

In addition to analyzing specific active-site residues, we compared the total active-site sequence composition of 25 ancestral nitrogenases (inferred across Tree-1–5) and all 284 extant nitrogenase sequences used for phylogenetic reconstruction. Our analysis shows clear distinctions between the active-site compositions of extant Mo- versus V-/Fe-nitrogenases (**Figure 6a**). Specifically, the mean identity between Vnf/Anf and Nif-I/Nif-II is ~50%, as compared with the mean identity between Vnf and Anf (~71%) and between Nif-I and Nif-II (~77%). The difference between Vnf/Anf and Nif-I/Nif-II sequences is not seen as distinctly when comparing total HDK sequences (**Figure 6b**), suggesting that the active-site sequence differences in particular may be better indicators of metal dependence than whole sequence differences due to greater catalytic tuning toward the cofactor.



505

506 **Figure 6.** Active-site and full HDK sequence comparisons between extant and ancestral
507 nitrogenases. (a) Active-site sequence identities of ancestral and extant nitrogenases. Active-site
508 residues include 30 amino acids positioned within 5 Å of the active-site cofactor. (b) Total HDK
509 sequence identities of ancestral and extant nitrogenases. Percentage identity values have been
510 averaged within each field of comparison. All 25 maximum likelihood ancestors and 284 extant
511 sequences included in this study were used for sequence identity calculation. Phylogeny coloring
512 is as follows: Clfx, yellow; F-Mc, green; Mb-Mc, purple; Anf/Vnf, red; Nif-I/-II, blue.

513

514 Because we observed active-site sequence distinctions between extant Vnf/Anf and Nif-I/Nif-II
515 nitrogenases, we compared active-site sequences of ancestral versus extant nitrogenases to provide
516 clues regarding ancestral metal dependence. Nearly all ancestral nitrogenases, including those
517 inferred for the oldest ancestral nodes, share greater active-site identity with Mo-nitrogenases than
518 with V-/Fe-only-nitrogenases (**Figure 6a**). Mean active-site sequence identities between AncB–
519 AncE and Anf/Vnf nitrogenases range between ~50 and 63%, whereas those between AncB–AncE
520 and Nif-I/Nif-II nitrogenases range between ~69 and 87%. An exception is AncA (ancestral to Vnf

521 and Anf), which has higher mean identity to Anf/Vnf nitrogenases (~85%) than to Nif-I/Nif-II
522 nitrogenases (~51%). Because active-site sequence identity can reliably differentiate extant Mo-
523 from V-/Fe-nitrogenases, the resemblance of most ancestral active sites to those of Mo-
524 nitrogenases is suggestive of Mo-dependence.
525

526 **4 DISCUSSION**

527 Nitrogenase mediates the reduction of N_2 to NH_3, a key step in nitrogen fixation (Anbar & Knoll,
528 2002; Canfield et al., 2010; Falkowski, 1997). The metal dependence of nitrogenase, which
529 impacts both catalytic properties (Eady, 1996; Harris, Yang, et al., 2018; Y. Hu et al., 2011; Lee
530 et al., 2018; Rebelein et al., 2018; Zheng et al., 2018) and ecological distribution (McRose, Zhang,
531 Kraepiel, & Morel, 2017; Zhang et al., 2016), suggests a potential role for marine geochemical
532 constraints on its evolution (Anbar & Knoll, 2002; Boyd, Hamilton, et al., 2011; Canfield et al.,
533 2010; Raymond et al., 2004). Thus, understanding ancestral nitrogenase metal dependence can
534 help resolve the early history of biological nitrogen fixation, and, in a broader sense, the impact
535 that ancient metal availabilities have had on the evolution of biologically essential metabolisms
536 over Earth history (Anbar & Knoll, 2002; Moore, Jelen, Giovannelli, Raanan, & Falkowski, 2017).
537 Previous phylogenetic work has established that Mo-, V-, and Fe-nitrogenases, though genetically
538 distinct, are evolutionarily homologous (Boyd, Anbar, et al., 2011; Boyd, Hamilton, et al., 2011;
539 Raymond et al., 2004). Most recent phylogenetic analyses also indicate that V- and Fe-
540 nitrogenases are derived from Mo-nitrogenases, the latter having originated following the gene
541 duplication event that produced *nifE* and *nifN* (Boyd, Anbar, et al., 2011). However, the precise
542 trajectory of metal-binding evolution in the nitrogenase family is not completely known, and

discrepancies between current phylogenetics-based models and the geochemical record of nitrogen fixation remain (Stueken et al., 2015).

We used both phylogenetic reconstruction and ancestral sequence inference to explore these outstanding questions of early nitrogenase evolution and metal dependence. Though the tree reconstructions presented here are largely in congruence with previous phylogenetic analyses, certain topological differences, particularly with regard to basal uncharacterized nitrogenases lacking associated *nifE* and/or *nifN* genes, suggest important deviations from previous narratives of early metal dependence (Boyd, Anbar, et al., 2011; Boyd, Hamilton, et al., 2011; Boyd & Peters, 2013; Raymond et al., 2004). The reconstruction of ancestral nitrogenase sequences *in silico* provides the means to directly infer ancient metal-binding properties from molecular information.

4.1 Inferred metal dependence of ancestral nitrogenases

The active-site protein environment is known, in addition to the metal content of the cofactor, to contribute to the variable catalytic properties of different nitrogenase isozymes (Brigle et al., 1987; Christiansen et al., 2000; Fixen et al., 2016; Kim et al., 1995; Sarma et al., 2010; Z. Y. Yang et al., 2012). We therefore explored active-site features of ancestral nitrogenases for correlations with metal dependence.

We find that modeled active-site structural features are not informative for the inference of ancestral metal dependence. Pocket volumes of modeled extant nitrogenases do not appear to be strongly correlated with metal cofactor content. Problematically, the volume ranges of oldest ancestors (as well as those of uncharacterized nitrogenases) overlap with both V- and Mo-

566 nitrogenases (**Figure 5**). We acknowledge that these homology modeling results may not precisely
567 reflect true biological differences (for which a comprehensive analysis is not yet possible due the
568 limited availability of V- and Fe-nitrogenase structures). This is illustrated by the difference
569 between pocket volumes of published structures and those of homology models (e.g., *A. vinelandii*
570 PDB 3U7Q NifD structure pocket volume ≈ 994.750 and mean *A. vinelandii* NifD modeled
571 structure pocket volume ≈ 1186.412). However, it is not surprising that active-site pocket volume
572 does not predict metal dependence, given the probable similar sizes and structures of the FeMo-,
573 FeV-, and FeFe-cofactors (Eady, 1996; Krahn et al., 2002; Daniel Sippel et al., 2018; Spatzal et
574 al., 2011). These findings contrast with a previous study by McGlynn and coworkers (2012) in
575 which pocket volume comparisons were used to classify the Mo-dependence of uncharacterized
576 nitrogenases. In this previous study, the volume means for Mo-nitrogenases were sufficiently
577 distinct from V- and Fe-nitrogenases as to provide unambiguous classification of Mo-dependence.
578 Our analyses differ from this previous study in several ways: (1) we incorporated V-nitrogenase
579 structural templates (D. Sippel & Einsle, 2017; Daniel Sippel et al., 2018) for homology modeling,
580 not available at the time of the previous study (2) we modeled 20 extant sequences of known metal
581 dependence rather than 12 (3) we used the same explicit parameters for both homology modeling
582 and pocket volume calculation across all sequences (4) we included modeling replicates for pocket
583 volume calculation. Thus, an expanded modeling analysis appears to reduce the efficacy of the
584 pocket volume parameter for inference of ancestral (and uncharacterized) nitrogenase metal
585 dependence.
586
587 Unlike structure-based analyses, we find that active-site sequence features reliably differentiate
588 Mo-nitrogenases and V-/Fe-nitrogenases. Regarding AncA (ancestral to V- and Fe-nitrogenases),

589 we identified specific active-site residues that have previously been suggested to interact with a
590 proposed carbonate ligand unique to the FeV-cofactor. These residues form a ^{355}TGGPRL^{360} loop
591 conserved only among V-nitrogenases and homologous to ^{355}IGGLRP^{360} in *A. vinelandii* NifD (D.
592 Sippel & Einsle, 2017). This substitution of Thr-355 for Ile-355, as well as the exchange of Leu
593 and Pro positions may permit the inclusion of the FeV-cofactor carbonate ligand by VnfD that is
594 not possible by the NifD protein (D. Sippel & Einsle, 2017). Thr-355 and Pro-358 are unique to
595 V-nitrogenases, and Leu-360 is unique to V- and Fe-nitrogenases. All AncA sequences conserve
596 the ^{355}TGGPRL^{360} residue loop capable of accommodating the FeV-cofactor. Furthermore, AncA
597 sequences generally exhibit greater numbers of residues unique to V-nitrogenases than those
598 unique to Fe-nitrogenases, and the mean identity of AncA active-site sequences is highest for V-
599 nitrogenases (**Figure 6a**). Together, these observations suggest that AncA is V-dependent.
600
601 Our comparisons of ancestral and extant sequence features indicate that the active-sites of oldest
602 nitrogenase ancestors (AncB–AncE) resemble those of extant Mo-nitrogenases more than V- or
603 Fe-nitrogenases (**Figure 6a**). This observation is particularly significant given that these same
604 patterns are not observed across the total HDK sequence (**Figure 6b**). Specifically, this
605 discrepancy supports the notion that the nitrogenase active-site has been tuned to the catalytic
606 properties of each metal cofactor over its evolutionary history (Harris, Yang, et al., 2018), and that
607 this tuning has manifested in active-site sequence differences that stand apart from baseline
608 phylogenetic distance. Though we are not able to identify specific residues that may functionally
609 relate to metal dependence (as with AncA), the resemblance of the early ancestral nitrogenase
610 active site to those of Mo-nitrogenases is highly suggestive of ancient Mo-dependence.
611

612 **4.2 A proposed model for the evolution of nitrogenase metal dependence over geologic time**

613 Despite the lack of information provided by structural analyses, the active-site sequence features
614 of oldest ancestral nitrogenases (i.e., AncB–E) support the inference of early Mo-dependence. The
615 observation that ancestral AncC and AncD active sites in particular most resemble those of extant
616 Mo-nitrogenases is at odds with the phylogenetic distribution of *nifE* and *nifN* genes, which
617 suggest that early-branching uncharacterized Clfx and F-Mc taxa (for which AncC and AncD are
618 ancestral) are not capable of assembling the FeMo-cofactor. The placement of Clfx and F-Mc
619 clades in our analyses differs from previous phylogenetic reconstructions. For example, the
620 phylogenetic tree presented by Boyd and coworkers nests Clfx and F-Mc clades within Mo-
621 nitrogenases, which notably branch more recently than V-, Fe-, and Mb-Mc Mo-nitrogenases
622 (Boyd, Hamilton, et al., 2011). A subsequently published topology is more similar to the tree
623 presented here, though lacking in Clfx sequences (Boyd, Costas, Hamilton, Mus, & Peters, 2015;
624 Boyd & Peters, 2013). It is possible that the larger sequence dataset used here has refined the
625 placement of these uncharacterized clades, which is supported by our analyses with an expanded
626 outgroup that maintains the positions of Clfx and F-Mc sequences (Appendix S2). Given the well-
627 supported placement of these uncharacterized clades in our reconstruction, we find that the
628 presence of *nifE* and *nifN* genes decreases stepwise with divergence age within the
629 uncharacterized/V-/Fe-nitrogenase lineage: Mb-Mc taxa, most recently branched, have both
630 *nifEN,* most F-Mc taxa only have *nifE*, and Clfx taxa, earliest branched, have neither. One may
631 thus parsimoniously conclude that uncharacterized nitrogenase AncC–D ancestors similarly
632 lacked the genetic requirements for FeMo-cofactor synthesis.

633

At least two possible models might resolve this discrepancy: (1) The last common nitrogenase ancestor represented in our tree was in fact hosted by an organism that possessed *nifEN* genes and was able to synthesize the FeMo-cofactor. However, this capability was lost in both Clfx and F-Mc taxa, but retained for all ancestors (AncA–E) as well as for Nif-I, Nif-II, and Mb-Mc taxa. The resemblance of oldest ancestral active sites (AncB–E) to those of Mo-nitrogenases indicate ancient FeMo-cofactor dependence, and the same resemblance of Clfx and F-Mc nitrogenases to Mo-nitrogenases may be inherited from these ancestors. (2) The last common nitrogenase ancestor was hosted by an organism that did not possess *nifEN* genes and was incapable of synthesizing the FeMo-cofactor. The presence of *nifEN* genes increased stepwise through the uncharacterized/V-/Fe-nitrogenase lineage, until the full FeMo-cofactor pathway was completed for AncB nitrogenase at the earliest, but at least in Mb-Mc nitrogenases (*nifEN* genes may have been transferred between ancestors of Mb-Mc, Nif-I, and Nif-II taxa). However, the resemblance of ancestral nitrogenase sequences to those of extant Mo-nitrogenases may evidence a currently unknown, alternative pathway for Mo-usage still present in extant Clfx and F-Mc taxa. This alternative, and possibly unrefined pathway for Mo-usage, may represent a transition state between ancient Mo-independence and full FeMo-cofactor usage.

We prefer the second model for several reasons:

(1) It is challenging to envision a scenario in which the FeMo-cofactor biosynthetic pathway would be lost in Clfx and F-Mc nitrogenases that appear to otherwise be capable of nitrogen fixation (Dekas et al., 2009; Keppen et al., 1989; Kuznetsov et al., 2011; Mehta & Baross, 2006). Mo-nitrogenases are far more efficient at reducing nitrogen than other isozymes (Eady, 1996; Harris

et al., 2019; Harris, Yang, et al., 2018), and the majority of all extant nitrogenases are Mo-dependent across both anoxic and oxic environments (Boyd et al., 2015; Mus, Colman, Peters, & Boyd, 2019; Raymond et al., 2004). Even those organisms that have additional V- or Fe-nitrogenases still retain and preferentially express Mo-nitrogenases (Boyd, Anbar, et al., 2011; Boyd, Hamilton, et al., 2011; Dos Santos et al., 2012; Hamilton et al., 2011; Raymond et al., 2004). Aside from Clfx and F-Mc clades, no other nitrogenases are known to lack associated *nifEN* genes.

(2) It has previously been proposed that early nitrogenases may have been capable of reducing nitrogen prior to the origin of *nifEN,* and thus of the FeMo-cofactor (Boyd, Hamilton, et al., 2011; Boyd & Peters, 2013; Mus et al., 2019; Soboh, Boyd, Zhao, Peters, & Rubio, 2010). Prior to the evolution of Mo-usage, an ancient Mo-independent nitrogenase may have been capable of — perhaps inefficiently — reducing nitrogen by a cofactor resembling the Fe-S-C cluster assembled by NifB, which constitutes the biosynthetic precursor to the FeMo-cofactor (Boyd & Peters, 2013; Mus et al., 2019; Soboh et al., 2010). Though our sequence analyses cannot assess ancestral nitrogenase dependence for a NifB-cofactor, it is likely that the NifB-cofactor resembles the structure and composition of the FeFe-cofactor, excepting homocitrate (Harris, Lukoyanov, et al., 2018). The greater similarity of AncC–D active sites to those of Mo-nitrogenases than Fe-nitrogenases likely suggests ancestral dependence on a cofactor incorporating Mo rather than only Fe. It is possible that, lacking NifEN, an alternative pathway for Mo-usage may have acted as a transition state between Mo-independence and full FeMo-cofactor usage. It is thus reasonable to speculate that this transition state of Mo-usage may be exhibited by AncC–D ancestors for which the lack of one or both *nifEN* genes might be parsimoniously inferred.

(3) Given greater catalytic efficiency of Mo-nitrogenases, even an unrefined pathway for Mo-usage may have been of selective advantage if Mo was transiently available prior to ~2.3–2.5 Ga Earth surface oxygenation (Anbar et al., 2007; Anbar & Knoll, 2002; Canfield et al., 2010; Lyons et al., 2014; Raymond et al., 2004). An early generalist behavior with regard to metal usage has previously been proposed (Raymond et al., 2004); such limited and possibly non-specific Mo-usage early in nitrogenase history would have provided an evolutionary pathway for increased Mo-dependence via the addition of NifEN biosynthetic components.

(4) ~3.2-Ga nitrogen isotopic signatures have been interpreted to reflect ancient Mo-dependent nitrogenase activity, due to similarities in isotopic fractionation by extant Mo-nitrogenases (Stueken et al., 2015). However, this interpretation conflicts with age estimates of both *nifEN* (and the FeMo-cofactor) (Boyd, Anbar, et al., 2011) and of earliest marine Mo availability (Anbar et al., 2007; Lyons et al., 2014). It is possible that alternative Mo-usage in early nitrogenases may have resulted in similar isotopic fractionations. Such possibilities may resolve this discrepancy and remain to be tested.

The preferred model of an ancestral, alternative pathway for Mo-usage, mapped to our phylogenetic reconstruction and inferred ancestors, is illustrated in **Figure 7**. In the first stage, represented by AncD and extant Clfx nitrogenases, nitrogen fixation is achieved by a possible alternative pathway for Mo-usage in the absence of NifEN, incorporating an unknown "proto-cofactor" (**Figure 7a**). Such unrefined Mo-usage is of advantage during transient Mo availabilities prior to ~2.3–2.5 Ga Earth surface oxygenation (Anbar et al., 2007). In the second stage, represented by AncC and extant F-Mc nitrogenases, a gene duplication event forms NifE, though

703 potentially not sufficient for FeMo-cofactor synthesis (**Figure 7b**). Alternative Mo-usage may still
704 be possible. In the third stage, represented by AncB, AncE, and extant Mb-Mc nitrogenases, a
705 subsequent gene duplication event forms NifN (**Figure 7c**). Together, NifE and NifN are capable
706 of synthesizing the FeMo-cofactor, resulting in canonical Mo-dependence. In the fourth stage,
707 gene duplication of *nif* results in, first, *vnf*, followed by *anf* (**Figure 7d**). These V- and Fe-
708 dependent enzymes rely on NifBEN biosynthetic components. Though the greater efficiency of
709 Mo-nitrogenase results in widespread diversification, V- and Fe-nitrogenases provide selective
710 advantage in microenvironments deficient in Mo. This model, built from the phylogenetic and
711 ancestral sequence inferences provided here, as well as from decades of previous geobiological
712 investigations of nitrogenase evolution, helps resolve outstanding questions regarding ancient
713 metal dependence, and, importantly, provides testable hypotheses for future investigations. Such
714 investigations may, for example, seek to clarify the nitrogen capability and metal dependence of
715 uncharacterized Clfx and F-Mc nitrogenases, as well as experimentally resurrect and characterize
716 ancestral nitrogenase sequences in the laboratory.
717

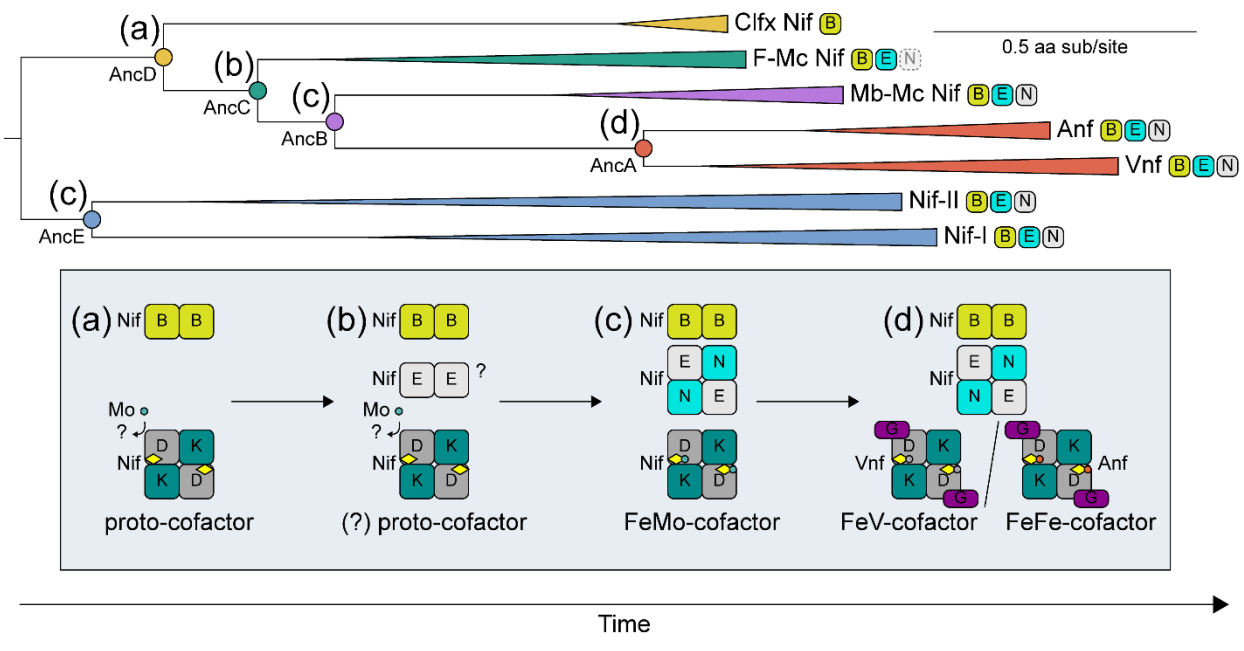


Figure 7. Proposed model for the evolution of nitrogenase metal specificity from an ancestral, alternative pathway for Mo-usage. Full description of stages (a)–(d) in the evolution of nitrogenase metal dependence is provided in the main text (Section 4.2). Possible alternative pathway for nitrogenase Mo-usage indicated in stages (a)–(b). Phase (a)–(d) are also mapped to analyzed ancestors within the nitrogenase phylogeny. Presence of NifB, NifE, and NifN in represented taxa indicated next to each clade of the phylogeny (F-Mc clade has only one species that harbors NifN).

4.3 Potential association of early nitrogen fixation with phototrophic metabolism

The relatively early-branching position of Clfx nitrogenases, hosted by strains in class Chloroflexi (green non-sulfur bacteria), suggests that biological nitrogen fixation may have had early associations with anoxygenic phototrophy. It has previously been proposed that biological nitrogen fixation originated in hydrogenotrophic methanogens due to the basal positioning of Mb-Mc (Methanobacteria and Methanococcus) or F-Mc (Firmicutes and Methanococcus) sequences in phylogenetic analyses of catalytic (i.e., NifHDK) and biosynthetic (i.e., NifBEN) nitrogenase

733 sequences (Boyd, Anbar, et al., 2011; Boyd, Hamilton, et al., 2011; Boyd & Peters, 2013;
734 Raymond et al., 2004). By contrast, in our phylogenetic reconstruction, Clfx NifHDK sequences
735 diverge earlier than both F-Mc and Mb-Mc sequences, with at least one represented strain having
736 been observed to fix nitrogen (Keppen et al., 1989; Kuznetsov et al., 2011). An early association
737 of biological nitrogen fixation with anoxygenic phototrophy would not be unexpected given the
738 well-studied homology between nitrogenases and dark-operative protochlorophyllide
739 oxidoreductases (Bch/ChlLNB), enzymes that catalyze part of the bacteriochlorophyll biosynthetic
740 pathway(Y. Hu & Ribbe, 2015; Moser & Brocker, 2011; Raymond et al., 2004). Furthermore,
741 Chloroflexi strains that possess uncharacterized Clfx nitrogenases also harbor BchLNB genes.
742 Though the phylogeny presented here may still be consistent with the origin of the full FeMo-
743 cofactor biosynthetic pathway (NifHDKBEN) in early methanogens (Boyd, Anbar, et al., 2011;
744 Boyd, Hamilton, et al., 2011; Boyd & Peters, 2013; Raymond et al., 2004), the basal position of
745 Clfx sequences may evidence an earlier metabolic context for biological nitrogen fixation prior to
746 the origin of the FeMo-cofactor. Future investigations of nitrogen fixation by other phototrophic
747 Chloroflexi strains and additional taxa that harbor basal uncharacterized nitrogenases may resolve
748 these aspects of nitrogenase origins.
749

750 **CONCLUSION**

751 We reconstructed the phylogenetic history of nitrogenase proteins, as well as inferred ancestral
752 nitrogenase sequences, in order to explore the evolutionary trajectory of nitrogenase metal
753 dependence. We find that, whereas modeled structural features of ancestral nitrogenases do not
754 offer conclusive indications of ancient metal usage, active-site sequence features of ancestors most
755 resemble those of extant Mo-nitrogenases. The absence of associated cofactor biosynthesis

proteins, considered necessary for FeMo-cofactor assembly, in several early-branching uncharacterized nitrogenase lineages evidences a possible alternative pathway for Mo-usage. We speculate that this alternative pathway may have preceded the evolution of the FeMo-cofactor and may today be present in extant uncharacterized nitrogenases, and we propose a model wherein canonical Mo-usage evolved via the stepwise introduction of FeMo-cofactor biosynthetic components following the divergence of more basal uncharacterized lineages. V-nitrogenases subsequently diversified, followed by Fe-nitrogenases, in agreement with previous phylogenetic inferences that Mo-dependence evolved first(Boyd, Hamilton, et al., 2011). This model helps to reconcile phylogenetic and geobiological explanations of nitrogenase evolution(Anbar & Knoll, 2002; Boyd, Anbar, et al., 2011; Boyd, Hamilton, et al., 2011). Future studies, particularly those that integrate experimental assessments of laboratory-resurrected ancestral nitrogenases, may continue to refine our understanding of nitrogenase and environmental coevolution.

ACKNOWLEDGEMENTS

This work was supported by a NASA Astrobiology Postdoctoral Fellowship (Garcia), the Harvard Origins of Life Initiative (Kacar, McShea), the National Science Foundation (Kacar, #1724090) and the John Templeton Foundation (Kacar, #58562 and #61239). We thank Andrew Knoll for constructive feedback.

REFERENCES

Aadland, K., Pugh, C., & Kolaczkowski, B. (2019). High-Throughput Reconstruction of Ancestral Protein Sequence, Structure, and Molecular Function. *Methods Mol Biol, 1851*, 135-170. doi:10.1007/978-1-4939-8736-8_8

Allen, R. M., Chatterjee, R., Ludden, P. W., & Shah, V. K. (1995). Incorporation of iron and sulfur from NifB cofactor into the iron-molybdenum cofactor of dinitrogenase. *J Biol Chem, 270*(45), 26890-26896. doi:10.1074/jbc.270.45.26890

Anbar, A. D., Duan, Y., Lyons, T. W., Arnold, G. L., Kendall, B., Creaser, R. A., . . . Buick, R. (2007). A whiff of oxygen before the great oxidation event? *Science, 317*(5846), 1903-1906. doi:10.1126/science.1140325

Anbar, A. D., & Knoll, A. H. (2002). Proterozoic ocean chemistry and evolution: a bioinorganic bridge? *Science, 297*(5584), 1137-1142. doi:10.1126/science.1069651

Anisimova, M., & Gascuel, O. (2006). Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst Biol, 55*(4), 539-552. doi:10.1080/10635150600755453

Benner, S. A., Sassi, S. O., & Gaucher, E. A. (2007). Molecular paleoscience: systems biology from the past. *Adv Enzymol Relat Areas Mol Biol, 75*, 1-132, xi.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., . . . Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Res, 28*(1), 235-242.

Boyd, E. S., Anbar, A. D., Miller, S., Hamilton, T. L., Lavin, M., & Peters, J. W. (2011). A late methanogen origin for molybdenum-dependent nitrogenase. *Geobiology, 9*(3), 221-232. doi:10.1111/j.1472-4669.2011.00278.x

Boyd, E. S., Costas, A. M., Hamilton, T. L., Mus, F., & Peters, J. W. (2015). Evolution of molybdenum nitrogenase during the transition from anaerobic to aerobic metabolism. *J Bacteriol, 197*(9), 1690-1699. doi:10.1128/JB.02611-14

Boyd, E. S., Hamilton, T. L., & Peters, J. W. (2011). An alternative path for the evolution of biological nitrogen fixation. *Front Microbiol, 2*, 205. doi:10.3389/fmicb.2011.00205

Boyd, E. S., & Peters, J. W. (2013). New insights into the evolutionary history of biological nitrogen fixation. *Front Microbiol, 4*, 201. doi:10.3389/fmicb.2013.00201

Brigle, K. E., Setterquist, R. A., Dean, D. R., Cantwell, J. S., Weiss, M. C., & Newton, W. E. (1987). Site-directed mutagenesis of the nitrogenase MoFe protein of Azotobacter vinelandii. *Proc Natl Acad Sci U S A, 84*(20), 7066-7069.

Bulen, W. A., & LeComte, J. R. (1966). The nitrogenase system from Azotobacter: two-enzyme requirement for N2 reduction, ATP-dependent H2 evolution, and ATP hydrolysis. *Proc Natl Acad Sci U S A, 56*(3), 979-986. doi:10.1073/pnas.56.3.979

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics, 10*, 421. doi:10.1186/1471-2105-10-421

Canfield, D. E., Glazer, A. N., & Falkowski, P. G. (2010). The evolution and future of Earth's nitrogen cycle. *Science, 330*(6001), 192-196. doi:10.1126/science.1186120

Chen, I. A., Chu, K., Palaniappan, K., Pillay, M., Ratner, A., Huang, J., . . . Kyrpides, N. C. (2019). IMG/M v.5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes. *Nucleic Acids Res, 47*(D1), D666-D677. doi:10.1093/nar/gky901

Christiansen, J., Cash, V. L., Seefeldt, L. C., & Dean, D. R. (2000). Isolation and characterization of an acetylene-resistant nitrogenase. *J Biol Chem, 275*(15), 11459-11464.

Curatti, L., Hernandez, J. A., Igarashi, R. Y., Soboh, B., Zhao, D., & Rubio, L. M. (2007). In vitro synthesis of the iron-molybdenum cofactor of nitrogenase from iron, sulfur, molybdenum, and homocitrate using purified proteins. *Proc Natl Acad Sci U S A, 104*(45), 17626-17631. doi:10.1073/pnas.0703050104

Dekas, A. E., Poretsky, R. S., & Orphan, V. J. (2009). Deep-sea archaea fix and share nitrogen in methane-consuming microbial consortia. *Science, 326*(5951), 422-426. doi:10.1126/science.1178223

Dos Santos, P. C., Fang, Z., Mason, S. W., Setubal, J. C., & Dixon, R. (2012). Distribution of nitrogen fixation and nitrogenase-like sequences amongst microbial genomes. *BMC Genomics, 13*, 162. doi:10.1186/1471-2164-13-162

Durrant, J. D., Votapka, L., Sorensen, J., & Amaro, R. E. (2014). POVME 2.0: An Enhanced Tool for Determining Pocket Shape and Volume Characteristics. *J Chem Theory Comput, 10*(11), 5047-5056. doi:10.1021/ct500381c

Eady, R. R. (1996). Structure-function relationships of alternative nitrogenases. *Chemical Reviews, 96*(7), 3013-3030. doi:10.1021/cr950057h

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res, 32*(5), 1792-1797. doi:10.1093/nar/gkh340

Falkowski, P. G. (1997). Evolution of the nitrogen cycle and its influence on the biological sequestration of CO2 in the ocean. *Nature, 387*(6630), 272-275. doi:10.1038/387272a0

Fixen, K. R., Zheng, Y., Harris, D. F., Shaw, S., Yang, Z. Y., Dean, D. R., . . . Harwood, C. S. (2016). Light-driven carbon dioxide reduction to methane by nitrogenase in a photosynthetic bacterium. *Proc Natl Acad Sci U S A, 113*(36), 10163-10167. doi:10.1073/pnas.1611043113

Garcia, A. K., & Kacar, B. (2019). How to resurrect ancestral proteins as proxies for ancient biogeochemistry. *Free Radic Biol Med*. doi:10.1016/j.freeradbiomed.2019.03.033

Gomez-Fernandez, B. J., Garcia-Ruiz, E., Martin-Diaz, J., Gomez de Santos, P., Santos-Moriano, P., Plou, F. J., . . . Alcalde, M. (2018). Directed -in vitro- evolution of Precambrian and extant Rubiscos. *Sci Rep, 8*(1), 5532. doi:10.1038/s41598-018-23869-3

Hageman, R. V., & Burris, R. H. (1978). Nitrogenase and nitrogenase reductase associate and dissociate with each catalytic cycle. *Proc Natl Acad Sci U S A, 75*(6), 2699-2702. doi:10.1073/pnas.75.6.2699

Hales, B. J., Case, E. E., Morningstar, J. E., Dzeda, M. F., & Mauterer, L. A. (1986). Isolation of a new vanadium-containing nitrogenase from Azotobacter vinelandii. *Biochemistry, 25*(23), 7251-7255. doi:10.1021/bi00371a001

Hamilton, T. L., Ludwig, M., Dixon, R., Boyd, E. S., Dos Santos, P. C., Setubal, J. C., . . . Peters, J. W. (2011). Transcriptional profiling of nitrogen fixation in Azotobacter vinelandii. *J Bacteriol, 193*(17), 4477-4486. doi:10.1128/JB.05099-11

Hanson-Smith, V., & Johnson, A. (2016). PhyloBot: A Web Portal for Automated Phylogenetics, Ancestral Sequence Reconstruction, and Exploration of Mutational Trajectories. *PLoS Comput Biol, 12*(7), e1004976. doi:10.1371/journal.pcbi.1004976

Hanson-Smith, V., Kolaczkowski, B., & Thornton, J. W. (2010). Robustness of ancestral sequence reconstruction to phylogenetic uncertainty. *Mol Biol Evol, 27*(9), 1988-1999. doi:10.1093/molbev/msq081

Harris, D. F., Lukoyanov, D., Kallas, H., Trncik, C., Yang, Z., Compton, P., . . . Seefeldt, L. C. (2019). Mo-, V-, and Fe-nitrogenases use a universal eight-electron reductive-elimination mechanism to achieve N_2 reduction. *Biochemistry*.

Harris, D. F., Lukoyanov, D. A., Shaw, S., Compton, P., Tokmina-Lukaszewska, M., Bothner, B., . . . Seefeldt, L. C. (2018). Mechanism of N2 Reduction Catalyzed by Fe-Nitrogenase Involves Reductive Elimination of H2. *Biochemistry, 57*(5), 701-710. doi:10.1021/acs.biochem.7b01142

Harris, D. F., Yang, Z. Y., Dean, D. R., Seefeldt, L. C., & Hoffman, B. M. (2018). Kinetic Understanding of N2 Reduction versus H2 Evolution at the E4(4H) Janus State in the Three Nitrogenases. *Biochemistry, 57*(39), 5706-5714. doi:10.1021/acs.biochem.8b00784

Hoffman, B. M., Lukoyanov, D., Yang, Z. Y., Dean, D. R., & Seefeldt, L. C. (2014). Mechanism of nitrogen fixation by nitrogenase: the next stage. *Chem Rev, 114*(8), 4041-4062. doi:10.1021/cr400641x

Hu, B., Harris, D. F., Dean, D. R., Liu, T. L., Yang, Z. Y., & Seefeldt, L. C. (2018). Electrocatalytic CO_2 reduction catalyzed by nitrogenase MoFe and FeFe proteins. *Bioelectrochemistry, 120*, 104-109. doi:10.1016/j.bioelechem.2017.12.002

Hu, Y., Lee, C. C., & Ribbe, M. W. (2011). Extending the Carbon Chain: Hydrocarbon Formation Catalyzed by Vanadium/Molybdenum Nitrogenases. *Science, 333*(6043), 753-755. doi:10.1126/science.1206883

Hu, Y., & Ribbe, M. W. (2011). Biosynthesis of Nitrogenase FeMoco. *Coord Chem Rev, 255*(9-10), 1218-1224. doi:10.1016/j.ccr.2010.11.018

Hu, Y., & Ribbe, M. W. (2015). Nitrogenase and homologs. *J Biol Inorg Chem, 20*(2), 435-445. doi:10.1007/s00775-014-1225-3

Joerger, R. D., & Bishop, P. E. (1988). Bacterial alternative nitrogen fixation systems. *Crit Rev Microbiol, 16*(1), 1-14. doi:10.3109/10408418809104465

Kacar, B., Hanson-Smith, V., Adam, Z. R., & Boekelheide, N. (2017). Constraining the timing of the Great Oxidation Event within the Rubisco phylogenetic tree. *Geobiology, 15*(5), 628-640. doi:10.1111/gbi.12243

Kennedy, C., & Dean, D. (1992). The nifU, nifS and nifV gene products are required for activity of all three nitrogenases of Azotobacter vinelandii. *Mol Gen Genet, 231*(3), 494-498.

Keppen, O. I., Lebedeva, N. V., Troshina, O. Y., & Rodionov, Y. V. (1989). The nitrogenase activity of filamentous phototrophic green bacterium. *Mikrobiologiya, 58*, 520-521.

Kim, C. H., Newton, W. E., & Dean, D. R. (1995). Role of the MoFe protein alpha-subunit histidine-195 residue in FeMo-cofactor binding and nitrogenase catalysis. *Biochemistry, 34*(9), 2798-2808.

Krahn, E., Weiss, R., Krockel, M., Groppe, J., Henkel, G., Cramer, P., . . . Muller, A. (2002). The Fe-only nitrogenase from Rhodobacter capsulatus: identification of the cofactor, an unusual, high-nuclearity iron-sulfur cluster, by Fe K-edge EXAFS and 57Fe Mossbauer spectroscopy. *J Biol Inorg Chem, 7*(1-2), 37-45. doi:10.1007/s007750100263

Kuznetsov, B. B., Ivanovsky, R. N., Keppen, O. I., Sukhacheva, M. V., Bumazhkin, B. K., Patutina, E. O., . . . Skryabin, K. G. (2011). Draft genome sequence of the anoxygenic filamentous phototrophic bacterium Oscillochloris trichoides subsp. DG-6. *J Bacteriol, 193*(1), 321-322. doi:10.1128/JB.00931-10

Le, S. Q., & Gascuel, O. (2008). An improved general amino acid replacement matrix. *Mol Biol Evol, 25*(7), 1307-1320. doi:10.1093/molbev/msn067

Lee, C. C., Tanifuji, K., Newcomb, M., Liedtke, J., Hu, Y., & Ribbe, M. W. (2018). A Comparative Analysis of the CO-Reducing Activities of MoFe Proteins Containing Mo- and V-Nitrogenase Cofactors. *Chembiochem, 19*(7), 649-653. doi:10.1002/cbic.201800035

Liu, Y., Schmidt, B., & Maskell, D. L. (2010). MSAProbs: multiple sequence alignment based on pair hidden Markov models and partition function posterior probabilities. *Bioinformatics, 26*(16), 1958-1964. doi:10.1093/bioinformatics/btq338

Lyons, T. W., Reinhard, C. T., & Planavsky, N. J. (2014). The rise of oxygen in Earth's early ocean and atmosphere. *Nature, 506*(7488), 307-315. doi:10.1038/nature13068

MacKay, B. A., & Fryzuk, M. D. (2004). Dinitrogen coordination chemistry: on the biomimetic borderlands. *Chem Rev, 104*(2), 385-401. doi:10.1021/cr020610c

McGlynn, S. E., Boyd, E. S., Peters, J. W., & Orphan, V. J. (2012). Classifying the metal dependence of uncharacterized nitrogenases. *Front Microbiol, 3*, 419. doi:10.3389/fmicb.2012.00419

McRose, D. L., Zhang, X., Kraepiel, A. M., & Morel, F. M. (2017). Diversity and Activity of Alternative Nitrogenases in Sequenced Genomes and Coastal Environments. *Front Microbiol, 8*, 267. doi:10.3389/fmicb.2017.00267

Mehta, M. P., & Baross, J. A. (2006). Nitrogen fixation at 92 degrees C by a hydrothermal vent archaeon. *Science, 314*(5806), 1783-1786. doi:10.1126/science.1134772

Moore, E. K., Jelen, B. I., Giovannelli, D., Raanan, H., & Falkowski, P. G. (2017). Metal availability and the expanding network of microbial metabolisms in the Archaean eon. *Nature Geoscience, 10*(9), 629-636. doi:10.1038/ngeo3006

Moser, J., & Brocker, M. J. (2011). Enzymatic Systems with Homology to Nitrogenase. *Nitrogen Fixation: Methods and Protocols, 766*, 67-77. doi:10.1007/978-1-61779-194-9_5

Mus, F., Alleman, A. B., Pence, N., Seefeldt, L. C., & Peters, J. W. (2018). Exploring the alternatives of biological nitrogen fixation. *Metallomics, 10*(4), 523-538. doi:10.1039/c8mt00038g

Mus, F., Colman, D. R., Peters, J. W., & Boyd, E. S. (2019). Geobiological feedbacks, oxygen, and the evolution of nitrogenase. *Free Radic Biol Med*. doi:10.1016/j.freeradbiomed.2019.01.050

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., . . . Pruitt, K. D. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res, 44*(D1), D733-745. doi:10.1093/nar/gkv1189

Ormeño-Orrillo, E., Hungria, M., & Martinez-Romero, E. (2013). Dinitrogen-Fixing Prokaryotes. In *The Prokaryotes* (pp. 427-451).

Quang, L. S., Gascuel, O., & Lartillot, N. (2008). Empirical profile mixture models for phylogenetic reconstruction. *Bioinformatics, 24*(20), 2317-2323. doi:10.1093/bioinformatics/btn445

R Core Team. (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Raymond, J., Siefert, J. L., Staples, C. R., & Blankenship, R. E. (2004). The natural history of nitrogen fixation. *Mol Biol Evol, 21*(3), 541-554. doi:10.1093/molbev/msh047

Rebelein, J. G., Lee, C. C., Newcomb, M., Hu, Y., & Ribbe, M. W. (2018). Characterization of an M-Cluster-Substituted Nitrogenase VFe Protein. *MBio, 9*(2). doi:10.1128/mBio.00310-18

Roll, J. T., Shah, V. K., Dean, D. R., & Roberts, G. P. (1995). Characteristics of NIFNE in Azotobacter vinelandii strains. Implications for the synthesis of the iron-molybdenum cofactor of dinitrogenase. *J Biol Chem, 270*(9), 4432-4437. doi:10.1074/jbc.270.9.4432

Rubio, L. M., & Ludden, P. W. (2008). Biosynthesis of the iron-molybdenum cofactor of nitrogenase. *Annu Rev Microbiol, 62*, 93-111. doi:10.1146/annurev.micro.62.081307.162737

Sali, A., & Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol, 234*(3), 779-815. doi:10.1006/jmbi.1993.1626

Sarma, R., Barney, B. M., Keable, S., Dean, D. R., Seefeldt, L. C., & Peters, J. W. (2010). Insights into substrate binding at FeMo-cofactor in nitrogenase from the structure of an alpha-70(Ile) MoFe protein variant. *J Inorg Biochem, 104*(4), 385-389. doi:10.1016/j.jinorgbio.2009.11.009

Schmid, B., Einsle, O., Chiu, H. J., Willing, A., Yoshida, M., Howard, J. B., & Rees, D. C. (2002). Biochemical and structural characterization of the cross-linked complex of nitrogenase: comparison to the ADP-AlF4(-)-stabilized structure. *Biochemistry, 41*(52), 15557-15565.

Setubal, J. C., dos Santos, P., Goldman, B. S., Ertesvag, H., Espin, G., Rubio, L. M., . . . Wood, D. (2009). Genome sequence of Azotobacter vinelandii, an obligate aerobe specialized to support diverse anaerobic metabolic processes. *J Bacteriol, 191*(14), 4534-4545. doi:10.1128/JB.00504-09

Shah, V. K., Allen, J. R., Spangler, N. J., & Ludden, P. W. (1994). In vitro synthesis of the iron-molybdenum cofactor of nitrogenase. Purification and characterization of NifB cofactor, the product of NIFB protein. *J Biol Chem, 269*(2), 1154-1158.

Shah, V. K., Imperial, J., Ugalde, R. A., Ludden, P. W., & Brill, W. J. (1986). In vitro synthesis of the iron-molybdenum cofactor of nitrogenase. *Proc Natl Acad Sci U S A, 83*(6), 1636-1640. doi:10.1073/pnas.83.6.1636

Shih, P. M., Occhialini, A., Cameron, J. C., Andralojc, P. J., Parry, M. A., & Kerfeld, C. A. (2016). Biochemical characterization of predicted Precambrian RuBisCO. *Nat Commun, 7*, 10382. doi:10.1038/ncomms10382

Sippel, D., & Einsle, O. (2017). The structure of vanadium nitrogenase reveals an unusual bridging ligand. *Nat Chem Biol, 13*(9), 956-960. doi:10.1038/nchembio.2428

Sippel, D., Rohde, M., Netzer, J., Trncik, C., Gies, J., Grunau, K., . . . Einsle, O. (2018). A bound reaction intermediate sheds light on the mechanism of nitrogenase. *Science, 359*(6383), 1484-1489. doi:10.1126/science.aar2765

Soboh, B., Boyd, E. S., Zhao, D., Peters, J. W., & Rubio, L. M. (2010). Substrate specificity and evolutionary implications of a NifDK enzyme carrying NifB-co at its active site. *FEBS Lett, 584*(8), 1487-1492. doi:10.1016/j.febslet.2010.02.064

Spatzal, T., Aksoyoglu, M., Zhang, L., Andrade, S. L., Schleicher, E., Weber, S., . . . Einsle, O. (2011). Evidence for interstitial carbon in nitrogenase FeMo cofactor. *Science, 334*(6058), 940. doi:10.1126/science.1214025

St John, R. T., Johnston, H. M., Seidman, C., Garfinkel, D., Gordon, J. K., Shah, V. K., & Brill, W. J. (1975). Biochemistry and genetics of Klebsiella pneumoniae mutant strains unable to fix N2. *J Bacteriol, 121*(3), 759-765.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics, 30*(9), 1312-1313. doi:10.1093/bioinformatics/btu033

Stueken, E. E., Buick, R., Guy, B. M., & Koehler, M. C. (2015). Isotopic evidence for biological nitrogen fixation by molybdenum-nitrogenase from 3.2 Gyr. *Nature, 520*(7549), 666-669. doi:10.1038/nature14180

Tal, S., Chun, T. W., Gavini, N., & Burgess, B. K. (1991). The delta nifB (or delta nifE) FeMo cofactor-deficient MoFe protein is different from the delta nifH protein. *J Biol Chem, 266*(16), 10654-10657.

Thornton, J. W. (2004). Resurrecting ancient genes: experimental analysis of extinct molecules. *Nat Rev Genet, 5*(5), 366-375. doi:10.1038/nrg1324

Whelan, S., & Goldman, N. (2001). A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol, 18*(5), 691-699. doi:10.1093/oxfordjournals.molbev.a003851

Yang, Z. (1993). Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. *Mol Biol Evol, 10*(6), 1396-1401. doi:10.1093/oxfordjournals.molbev.a040082

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol, 24*(8), 1586-1591. doi:10.1093/molbev/msm088

Yang, Z. Y., Moure, V. R., Dean, D. R., & Seefeldt, L. C. (2012). Carbon dioxide reduction to methane and coupling with acetylene to form propylene catalyzed by remodeled nitrogenase. *Proc Natl Acad Sci U S A, 109*(48), 19644-19648. doi:10.1073/pnas.1213159109

Zhang, X., McRose, D. L., Darnajoux, R., Bellenger, J. P., Morel, F. M. M., & Kraepiel, A. M. L. (2016). Alternative nitrogenase activity in the environment and nitrogen cycle implications. *Biogeochemistry, 127*(2-3), 189-198. doi:10.1007/s10533-016-0188-6

Zheng, Y., Harris, D. F., Yu, Z., Fu, Y., Poudel, S., Ledbetter, R. N., . . . Harwood, C. S. (2018). A pathway for biological methane production using bacterial iron-only nitrogenase. *Nat Microbiol, 3*(3), 281-286. doi:10.1038/s41564-017-0091-5