Long title: Nitrogenase resurrection and the evolution of a singular enzymatic mechanism

Short title: Nitrogenase resurrection and evolution

Amanda K. Garcia[a], Derek F. Harris[b], Alex J. Rivier[a], Brooke M. Carruthers[a], Azul Pinochet-Barros[a], Lance C. Seefeldt[b], Betül Kaçar[a]*

[a] Department of Bacteriology, University of Wisconsin–Madison, Madison, WI 53706

[b] Department of Chemistry and Biochemistry, Utah State University, Logan, UT 84322

*To whom correspondence should be addressed: Betül Kaçar, [602 263 3622], bkacar@wisc.edu

Keywords: ancestral sequence reconstruction, evolutionary biochemistry, metabolic engineering, nitrogen fixation, nitrogenase

ABSTRACT

The planetary biosphere is powered by a suite of key metabolic innovations that emerged early in the history of life. However, it is unknown whether life has always followed the same set of strategies for performing these critical tasks. Today, microbes access atmospheric sources of bioessential nitrogen through the activities of just one family of enzymes, nitrogenases. Here, we show that the only dinitrogen reduction mechanism known to date is an ancient feature conserved from nitrogenase ancestors. We designed a paleomolecular engineering approach wherein ancestral nitrogenase genes were phylogenetically reconstructed and inserted into the genome of the diazotrophic bacterial model, *Azotobacter vinelandii,* enabling an integrated assessment of both *in vivo* functionality and purified nitrogenase biochemistry. Nitrogenase ancestors are active and robust to variable incorporation of one or more ancestral protein subunits. Further, we find that all ancestors exhibit the reversible enzymatic mechanism for dinitrogen reduction, specifically evidenced by hydrogen inhibition, that is also exhibited by extant *A. vinelandii* nitrogenase isozymes. Our results suggest that life may have been constrained in its sampling of protein sequence space to catalyze one of the most energetically challenging biochemical reactions in nature. The experimental framework established here is essential for probing how nitrogenase functionality has been shaped within a dynamic, cellular context to sustain a globally consequential metabolism.

IMPACT STATEMENT

The enzymatic mechanism for dinitrogen reduction is an ancient feature of nitrogenases that persisted over hundreds of millions of years.

INTRODUCTION

The evolutionary history of life on Earth has generated tremendous ecosystem diversity, the sum of which is orders of magnitude larger than that which exists at present [1]. Life's historical diversity provides a measure of its ability to solve adaptive problems within an integrated planetary system. Accessing these solutions requires a deeper understanding of the selective forces that have shaped the evolution of molecular-scale, metabolic innovations. However, the early histories of many of life's key metabolic pathways and the enzymes that catalyze them remain coarsely resolved.

Important efforts to advance understanding of early metabolic innovations have included phylogenetic inference [2, 3], systems-level network reconstructions [4], and the leveraging of extant or mutant biological models as proxies for their ancient counterparts [5-7]. However, methods to directly study these metabolic evolutionary histories across past environmental and cellular transitions remain underexplored. A unified, experimental strategy that integrates historical changes to enzymes, which serve as the primary interface between metabolism and environment, and clarifies their impact within specific cellular and physiochemical contexts is necessary. To address this, phylogenetic reconstructions of enzymes can be directly integrated within laboratory microbial model systems [8-10]. In this paleomolecular framework, predicted ancestral enzymes can be "resurrected" within a compatible host organism for functional characterization. These experimental systems can ultimately integrate multiple levels of historical analysis by interrogating critical features of ancient enzymes as well as dynamic interactions between enzymes, their broader metabolic networks, and the external environment.

The study of biological nitrogen fixation offers a promising testbed to thread these investigations of early metabolic evolution. Both phylogenetic and geological evidence [11-15] indicate that the origin of biological nitrogen fixation was a singular and ancient evolutionary event on which the modern biosphere has since been built. The only known nitrogen fixation pathway (compared to, for example, at least seven carbon-fixation pathways [16]) is catalyzed by an early evolved family of metalloenzymes called nitrogenases that reduce highly inert, atmospheric dinitrogen (N_2) to bioavailable ammonia (NH_3). The nitrogenase family comprises three isozymes that vary in their metal dependence (i.e., molybdenum, vanadium, and iron) and, in certain cases, all coexist within the same host organism [17]. Many diazotrophs depend on genetic strategies for coordinating the biosynthesis and expression of multiple nitrogenase isozymes and their respective

metalloclusters [18, 19], and, in oxic environments, protecting the oxygen-sensitive metalloclusters from degradation [20]. Thus, nitrogenase enzymes are a central component of a broader, co-evolving nitrogen fixation machinery. These features that create significant experimental challenges for nitrogen fixation engineering [21-23] nevertheless also make this metabolism an ideal candidate for systems-level, paleomolecular study.

How biological nitrogen fixation emerged and evolved under past environmental conditions is still poorly constrained relative to its importance in Earth's planetary and biological history. Because nitrogen has been a limiting nutrient over geological timescales [24, 25], nitrogenase has long been a key constituent of the expanding Earth biosphere. The impact of nitrogen limitation is underscored by human reliance on the industrial Haber-Bosch process, an energetically and environmentally costly workaround for nitrogen fertilizer production [26] designed to supplement a remarkable molecular innovation that biology has tinkered with for more than three billion years. How the structural domains and regulatory network of nitrogenase were recruited [27, 28] and under what selective pressures the metal dependence of nitrogenases were shaped [11, 29] remain open questions. Importantly, it is not known how the enzymatic mechanism for dinitrogen reduction has been tuned by both peptide and metallocluster to achieve one of the most difficult reactions in nature [30-32]. At the enzyme level, previous insights into nitrogenase sequence-function relationships have primarily derived from single or dual substitution studies. These have often yielded diminished or abolished nitrogenase activity [30, 31], though in certain cases improved reactivity toward alternate, industrially relevant substrates [30]. Despite illuminating key features of extant nitrogenase mechanism in select model organisms, the combination of detailed functional studies within an explicit evolutionary scheme has not previously been accomplished for the nitrogen fixation system.

Here, we seek guidance from the Earth's evolutionary past to reconstruct the history of the key metabolic enzyme, nitrogenase. We establish an evolutionary systems biology approach for the cellular- and molecular-level characterization of ancestral nitrogenases resurrected within the model diazotrophic bacterium, *Azotobacter vinelandii* (*A. vinelandii*). We find that variably replacing different protein subunits of the nitrogenase complex with inferred ancestral counterparts enables nitrogen fixation in *A. vinelandii*. Purified ancestral enzymes exhibit the specific N_2 reduction mechanism retained by their studied, extant counterparts, and maintain the same catalytic selectivity between N_2 and protons. Thus, the core strategy for biological nitrogen fixation is conserved across the investigated timeline. Our paleomolecular approach opens a new

109 route to study the ancient functionality and evolution of nitrogenases both deeper in its ancestry
110 and within the broader context of its supporting, cellular machinery.
111
112 **RESULTS**
113
114 **A resurrection strategy for ancestral nitrogenases**
115
116 We designed an engineering pipeline for the resurrection and experimental characterization of
117 ancestral nitrogenases (**Fig. 1*A***). In this scheme, phylogenetically inferred, ancestral nitrogenase
118 genes are synthesized and engineered into the genome of a modern diazotrophic bacterium,
119 enabling the assessment of *in vivo* nitrogenase activity and expression in parallel with biochemical
120 analysis of purified enzyme. The engineering and functional assessment of ancestral
121 nitrogenases required a suitable diazotrophic microbial host, owing to challenges associated with
122 nitrogenase heterologous expression [26]. We selected the obligately aerobic
123 gammaproteobacterium, *A. vinelandii* (strain "DJ"), an ideal experimental model due to its genetic
124 tractability and the availability of detailed studies on the genetics and biochemistry of its nitrogen
125 fixation machinery [33]. We specifically targeted the extant *A. vinelandii* molybdenum-dependent
126 ("Mo-") nitrogenase (hereafter referred to as wild-type, "WT"), which is the best studied isozyme
127 [30] relative to the *A. vinelandii* vanadium ("V-") and iron ("Fe-") dependent nitrogenases . The
128 WT Mo-nitrogenase complex comprises multiple subunits, NifH, NifD, and NifK (encoded by
129 *nifHDK* genes), which are arranged into two catalytic components: a NifH homodimer and a NifDK
130 heterotetramer (**Fig. 1*B***). During catalysis, both components transiently associate to transfer one
131 electron from NifH to NifDK and subsequently dissociate. Transferred electrons accumulate at the
132 active-site Mo-containing metallocluster ("FeMoco") housed within the NifD subunits for reduction
133 of N_2 substrate to NH_3.
134
135 To infer Mo-nitrogenase ancestors, we built a maximum-likelihood nitrogenase phylogeny from a
136 concatenated alignment of NifHDK amino acid sequences (**Fig. 2*A*; Fig. 2-Figure supplement**
137 **1**). The phylogeny contains 385 sets of homologs representative of known nitrogenase molecular
138 sequence diversity (including Mo-, V-, and Fe-nitrogenases), and is rooted by dark-operative
139 protochlorophyllide oxidoreductase proteins classified within the nitrogenase superfamily [34]. For
140 this study, we selected ancestors that fall within the direct evolutionary lineage of *A. vinelandii*
141 WT (**Fig. 2*B***), "Anc1" and "Anc2" (listed in order of increasing age), having ~90% and ~85% amino
142 acid sequence identity to WT across the full length of their concatenated NifHDK proteins,

143 respectively (**Fig. 2*C*; Supplementary File 1a**). A relatively conservative percentage identity
144 threshold was chosen based on prior studies benchmarking functional expression of ancestral
145 elongation factor proteins in *Escherichia coli* [35] and *Synechococcus elongatus* [10]. The high-
146 dimensional, nitrogenase protein sequence space occupied by both extant and ancestral
147 homologs is visualized in two-dimensions in **Fig. 2*D*** by machine-learning embeddings (see
148 **Materials and methods**). This analysis highlights the swath of sequence space targeted here,
149 as well as that made accessible by resurrection of nitrogenase ancestors more broadly. WT,
150 Anc1, and Anc2 lie within a Mo-nitrogenase clade (previously termed "Group I" [12]) that contains
151 homologs from diverse aerobic and facultatively anaerobic taxa, including proteobacteria and
152 cyanobacteria (**Fig. 2*A***). A maximum age constraint of ~2.5 Ga for Group I nitrogenases (and
153 thus for both Anc1 and Anc2) can be reasoned based on timing of the Great Oxidation Event [36]
154 and downstream emergence of aerobic taxa represented nearly exclusively within this clade.
155
156 Residue-level differences between ancestral and WT nitrogenases ("ancestral substitutions") are
157 broadly distributed along the length of each ancestral sequence (**Fig. 1*B*; Fig. 2-Figure**
158 **Supplement 2**). An ancestral substitution proximal to the active-site FeMoco metallocluster lies
159 within a loop considered important for FeMoco insertion [37] (NifD I355V; residue numbering from
160 WT)) and is observed across all targeted NifD ancestors (**Fig. 1*B***). Other ancestral substitutions
161 are notable for their location at relatively conserved residue sites (assessed by ConSurf [38]; **Fig.**
162 **2-Figure Supplement 3**; see **Materials and methods**) and/or within subunit interfaces, including
163 two at the NifD:NifK interface that are proximal to one another, F429Y (NifD) and R108K (NifK).
164 The Anc2 NifD protein contains five more ancestral substitutions at conserved sites than the
165 younger Anc1 NifD protein. In all studied ancestors, the C275 and H442 FeMoco ligands, as well
166 as other strictly conserved nitrogenase residues, are retained.
167
168 Phylogenetic analysis informs the compatibility of selected ancestors in extant microbial hosts.
169 Extant nitrogenases within the Group I nitrogenase clade (which include Anc1 and Anc2
170 descendants) are associated with numerous accessory genes likely recruited to optimize
171 synthesis and regulation of the oxygen-sensitive nitrogenase for aerobic or facultative
172 metabolisms [27]. For example, in addition to the structural *nifHDK* genes, *A. vinelandii* WT is
173 assembled and regulated with the help of >15 additional *nif* genes. Likewise, the extant
174 descendants of Anc1 and Anc2 are primarily aerobic or facultative proteobacteria and are thus
175 associated with higher complexity *nif* gene clusters (**Fig. 2*B***). We hypothesized that the likely
176 oxygen-tolerant, ancient proteobacterium harboring these ancestral nitrogenases were similar in

177 *nif* cluster complexity to extant *A. vinelandii*. Thus, we predicted that the *nif* accessory genes
178 present in *A. vinelandii* would support the functional expression of resurrected nitrogenase
179 ancestors.
180
181 To gauge the degree of compatibility between ancestral and extant nitrogenase proteins and
182 maximize the chance of recovering functional nitrogenase variants, we executed two parallel
183 genome engineering strategies. First, we constructed *A. vinelandii* strains harboring only the
184 ancestral *nifD* gene from both targeted ancestral nodes ("Anc1A" and "Anc2"), thereby expressing
185 "hybrid ancestral-WT" nitrogenase complexes (**Fig. 1*C***). This strategy is similar to *in vitro* "cross-
186 reaction" studies that have evaluated the compatibility of nitrogenase protein components from
187 differing host taxa [39]. Second, we constructed a strain harboring all Anc1 *nifHDK* genes,
188 expressing a fully ancestral nitrogenase complex ("Anc1B"; sequence reconstructed from a node
189 equivalent to Anc1A from an alternate phylogeny, see **Materials and methods**). *A. vinelandii*
190 strains were constructed by markerless genomic integration of ancestral nitrogenase genes, as
191 described in **Materials and methods**.
192
193 **Ancestral nitrogenases enable diazotrophic microbial growth**
194
195 All *A. vinelandii* constructs harboring ancestral genes enabled diazotrophic growth in
196 molybdenum-containing, nitrogen-free media. All strains had comparable doubling times to WT
197 during the exponential phase ($p > .05$; **Fig. 3*A*, 3*B***). The only significant difference among strains
198 was a ~14-h increase in the lag phase of strain Anc2 relative to WT, harboring the oldest
199 nitrogenase ancestor ($p \approx 2e\text{-}7$). We did not detect growth under the same conditions for a control
200 Δ*nifD* strain (DJ2278, see **Supplementary File 1c**). This result confirmed that the growth
201 observed for ancestral strains did not stem from leaky expression of the alternative, V- or Fe-
202 dependent nitrogen fixation genes in *A. vinelandii,* which were left intact.
203
204 An acetylene reduction assay was performed to measure cellular nitrogenase activity in
205 engineered strains. This assay quantifies the reduction rate of the non-physiological substrate
206 acetylene (C_2H_2) to ethylene (C_2H_4) [40], here normalized to total protein content. *A. vinelandii*
207 strains harboring only ancestral *nifD* (Anc1A, Anc2) exhibited mean C_2H_2 reduction rates of ~5 to
208 6 μmol C_2H_4/mg total protein/h, ~40-45% that of WT ($p \approx 6e\text{-}4$ and $p \approx 3e\text{-}4$, respectively) (**Fig.
209 3*C***). Strain Anc1B, harboring ancestral *nifHDK,* exhibited a mean acetylene reduction rate of ~9
210 μmol C_2H_4/mg total protein/h, ~70% that of WT ($p \approx 3e\text{-}2$).

211
212 The phenotypic variability we observed among engineered and WT *A. vinelandii* strains might
213 result both from differences in nitrogenase expression and nitrogenase activity. To provide
214 insights into these disparate effects, we quantified nitrogenase protein expression in engineered
215 strains by immunodetection of ancestral and WT Strep-tagged NifD proteins (the latter from strain
216 DJ2102, see **Supplementary File 1c**) and did not conclusively detect significant differences in
217 protein quantity relative to WT ($p > .05$; **Fig. 3*D***).
218
219 **Purified ancestral nitrogenases conserve extant N_2 reduction mechanism and efficiency**
220
221 Ancestral nitrogenase NifDK protein components were expressed and purified for biochemical
222 characterization. All ancestral NifDK proteins were assayed with WT NifH protein (ancestral NifH
223 proteins were not purified) for reduction of H^+, N_2, and C_2H_2. Ancestors were found to reduce all
224 three substrates *in vitro*, supporting the cellular-level evidence of ancestral nitrogenase activity
225 (**Fig. 4*A***).
226
227 We investigated whether the ancestral nitrogenases studied here would exhibit the general
228 mechanism for N_2 binding and reduction that has been observed for the studied, extant
229 nitrogenase isozymes of *A. vinelandii* (Mo, V, and Fe) [32, 41]. This mechanism involves the
230 accumulation of 4 electron/protons on the active-site cofactor as metal-bound hydrides,
231 generating the E_4(4H) state (**Fig. 4*B***). Once generated, N_2 can bind to the E_4(4H) state through a
232 reversible reductive elimination/oxidative addition (*re*/*oa*) mechanism, which results in the release
233 (*re*) of a single molecule of hydrogen gas (H_2). N_2 binding is reversible in the presence of sufficient
234 H_2, which displaces bound N_2 and results in reformation of E_4(4H) with two hydrides (*oa*). Thus,
235 a classic test of the (*re*/*oa*) mechanism is the ability of H_2 to inhibit N_2 reduction. We observed
236 that the reduction of N_2 to NH_3 for all nitrogenase ancestors was inhibited in the presence of H_2,
237 indicating that the ancestors follow the same mechanism of N_2 binding determined for extant
238 enzymes (**Fig. 4*C***).
239
240 In the event the E_4(4H) state fails to capture N_2, nitrogenases will simply produce H_2 from the
241 E_4(4H) state to generate the E_2(2H) state. The ratio of H_2 formed to N_2 reduced (H_2/N_2) can be
242 used as a measure of the efficiency of nitrogenases in using ATP and reducing equivalents for N_2
243 reduction. The stoichiometric minimum of the mechanism is $H_2/N_2 = 1$. Experimentally (under 1
244 atm N_2), a ratio of ~2 is seen for Mo-nitrogenase and ~5 and ~7 for V- and Fe-nitrogenase,

245 respectively [32]. H_2/N_2 values for all ancestors under 1 atm N_2 was ~2, similar to extant Mo-
246 nitrogenase (**Fig. 4*D***).
247
248 **DISCUSSION**
249
250 In this study, we leverage a new approach to investigate ancient nitrogen fixation by the
251 resurrection and functional assessment of ancestral nitrogenase enzymes. We demonstrate that
252 engineered *A. vinelandii* cells can reduce N_2 and C_2H_2 and exhibit diazotrophic growth rates
253 comparable to WT, though we observe that the oldest ancestor, Anc2, has a significantly longer
254 lag phase. Purified nitrogenase ancestors are active for reduction of H^+, N_2, and C_2H_2, while
255 maintaining the catalytic efficiency (described by the H_2/N_2 ratio) of WT enzymes. Our results also
256 show that ancestral N_2 reduction is inhibited by H_2, indicating an early emergence of the reductive-
257 elimination N_2-reduction mechanism preserved by characterized, extant nitrogenases [32, 41].
258 These properties are maintained despite substantial residue-level changes to the peripheral
259 nitrogenase structure (including relatively conserved sites), as well as a handful within the active-
260 site or protein-interface regions within the enzyme complex.
261
262 It is important to consider that the nitrogenase ancestors resurrected here represent hypotheses
263 regarding the true ancestral state. Uncertainty underlying ancestral reconstructions might derive
264 from incomplete extant molecular sequence data, as well as incorrect assumptions associated
265 with the implemented evolutionary models [8]. For example, a complicating feature of nitrogenase
266 evolution is that it has been shaped significantly by horizontal gene transfer [12, 15], which in
267 certain cases has led to different evolutionary trajectories of individual nitrogenase structural
268 genes. Specifically, certain H-subunit genes of the V-nitrogenase appear to have different
269 evolutionary histories relative to their DK components [12]. However, since our study targets a
270 Mo-nitrogenase lineage, we do not expect horizontal transfer to be a significant source of
271 uncertainty in our reconstructions.
272
273 The N_2-reduction activity of nitrogenase ancestors suggests that the required protein-protein
274 interactions—both between subunits that comprise the nitrogenase complex as well as those
275 required for nitrogenase assembly in *A. vinelandii*—and metallocluster interactions are sufficiently
276 maintained for primary function. Still, our results reveal the degree to which the organism-level
277 phenotype of host strains can be perturbed by varying both the number and age of ancestral
278 subunits. Importantly, these changes appear to impact phenotypic properties in complex ways,

representative of the type of cellular constraints on nitrogenase evolution that would be unobservable through an *in vitro* study alone. For example, we observed comparable growth characteristics of strains harboring single (Anc1A, ancestral NifD) versus multiple (Anc1B, ancestral NifHDK) ancestral subunits of equivalent age, whereas the lag phase of Anc2 hosting a single, older subunit (ancestral NifD) was increased. Here, growth is sensitive to older, ancestral substitutions in a single subunit while permissive of more recent ancestral substitutions across one or more subunits within the nitrogenase complex. However, a different pattern is observed across *in vivo* acetylene reduction rates. These are most negatively impacted relative to WT in strains with a single NifD ancestor (Anc1A, Anc2), whereas rates are more modestly decreased in a strain with a complete, ancestral NifHDK complex (Anc1A). These results suggest that ancestral subunits of equivalent age have greater compatibility and yield greater *in vivo* activity compared to subunits of disparate ages, perhaps owing to modified protein interactions within the nitrogenase complexes The discrepancy between *in vivo* activity and growth characteristics may also be attributable to impacted cellular processes external to the biochemical properties of the nitrogenase complex itself, and yet nevertheless vital in determining the overall fitness of the host organism. Finally, though we do not detect significant differences in ancestral protein expression here, it is possible that phenotypic outcomes of future reconstructions might be impacted by perturbed expression levels (e.g., [10, 42]). To what degree these expression levels are representative of the ancestral state and impact the phenotypic property of interest should be considered in future work.

That nitrogenase ancestors perform the reductive-elimination N_2-reduction mechanism—as the distantly related [11], extant Mo-, V-, and Fe-nitrogenases of *A. vinelandii* do today [32]—likely indicates that this enzymatic characteristic was set early in nitrogenase evolutionary history and sustained through significant past environmental change [36, 43, 44] and ecological diversification [27, 45]. It is possible that life's available strategies for achieving N_2 reduction may be fundamentally limited, and that a defining constraint of nitrogenase evolution has been the preservation of the same N_2 reduction mechanism across shifting selective pressures. For example, in the acquisition of V- and Fe-dependence from Mo-dependent ancestors [11], nitrogenases may have required substantial sequence and structural changes [46, 47] in order to facilitate reductive-elimination given a different active-site metallocluster. It is also possible that alternate strategies for biological nitrogen fixation evolved early in the history of life and were subsequently outcompeted, leaving no trace of their existence in extant microbial genomes. Why these alternate possibilities were evidently not explored by nature to the same degree remains an

open question, particularly given the several abiotic mechanisms for nitrogen fixation [48-50] and the multiple biological pathways for another, globally significant metabolism, carbon fixation [16]. Because our paleomolecular approach is ultimately informed by extant sequence data, it cannot directly evaluate extinct sequences that, for instance, due to contingency or entrenchment, did not persist and become preserved in extant microbial genomes. Nevertheless, evolutionarily informed studies of nitrogenase functionality that define the sequence-function space of this enzyme family will provide a foundation for laboratory efforts aimed toward exploring alternate scenarios. Future work that explores deeper into nitrogenase evolutionary history (and across extant and ancestral nitrogenase sequence space, as charted here (**Fig. 2*D***)) will clarify the degree of functional constraint exhibited by the nitrogenase family, both past and present.

CONCLUSION

Broadening the historical level of analysis beyond a single enzyme to the organism level is necessary to generate comprehensive insights into the evolutionary history and engineering potential of nitrogen fixation. Paleomolecular work that has expanded toward the systems-level investigation of early evolved, crucial metabolic pathways remains in its infancy, despite the potential for provocative connections between molecular-scale innovations and planetary history [8, 10]. Our results highlight the evolutionary conservation of a critical metabolic pathway that has shaped the biosphere over billions of years, as well as establish the tractability of leveraging phylogenetic models to carry out extensive, empirical manipulations of challenging enzymatic systems and their microbial hosts. Building on the empirical framework presented here will illuminate the evolutionary design principles behind ancient metabolic systems more broadly as well as leverage these histories to understand how key enzymes that allowed organisms to access nitrogen from the atmosphere evolved.

340 **MATERIALS AND METHODS**

341

Key Resources Table				
Reagent type (species) or resource	**Designation**	**Source or reference**	**Identifiers**	**Additional information**
strain, strain background (*Azotobacter vinelandii*)	DJ	DOI: 10.1128/JB.00504-09	n/a	Dennis Dean, Virginia Tech; Wild-type (WT); Nif+
genetic reagent (*Azotobacter vinelandii*)	DJ2102	DOI: 10.1016/bs.mie.2018.10.007	n/a	Dennis Dean, Virginia Tech; Strep-tagged WT NifD; Nif+
genetic reagent (*Azotobacter vinelandii*)	DJ2278	Other	n/a	Dennis Dean, Virginia Tech; Δ*nifD*::KanR; Nif-
genetic reagent (*Azotobacter vinelandii*)	DJ884	Other	n/a	Dennis Dean, Virginia Tech; *nifD*R187I mutant; Nif+(slow); overexpresses NifH
genetic reagent (*Azotobacter vinelandii*)	AK022	This paper	n/a	Δ*nifHDK*::KanR; Nif-

genetic reagent (*Azotobacter vinelandii*)	AK013	This paper	n/a	“Anc1A”; Δ*nifD*::*nifD*^{Anc1A}; Nif+
genetic reagent (*Azotobacter vinelandii*)	AK023	This paper	n/a	“Anc1B”; Δ*nifHDK*::*nifHDK*^{Anc1B}; Nif+
genetic reagent (*Azotobacter vinelandii*)	AK014	This paper	n/a	“Anc2”; Δ*nifD*::*nifD*^{Anc2}; Nif+
antibody	StrepMAB-Classic (Mouse monoclonal)	IBA Lifesciences	Cat# 2-1507-001, RRID: AB_513133	WB (1:5000)
recombinant DNA reagent	pAG25	This paper	n/a	KanR cassette (APH(3’)-I gene) + 400-bp *nifHDK* flanking homology regions, synthesized into XbaI/KpnI sites in pUC19; used to construct strain AK022 from DJ
recombinant DNA reagent	pAG13	This paper	n/a	*nifD*^{Anc1A} + 400-bp *nifD* flanking homology regions, synthesized into XbaI/KpnI sites in pUC19; used to construct strain Anc1A from AK022

recombinant DNA reagent	pAG19	This paper	n/a	*nifHDK*^{Anc1B} + 400-bp *nifHDK* flanking homology regions, synthesized into XbaI/KpnI sites in pUC19; used to construct strain Anc1B from AK022
recombinant DNA reagent	pAG14	This paper	n/a	*nifD*^{Anc2} + 400-bp *nifD* flanking homology regions, synthesized into XbaI/KpnI sites in pUC19; used to construct strain Anc2 from AK022
sequence-based reagent	306_nifH_F	This paper	PCR primers	GCCGAACGTTCAAGTGGAAA
sequence-based reagent	307_nifH_R	This paper	PCR primers	AGAGCCAATCTGCCCTGTC
sequence-based reagent	308_nifD_F	This paper	PCR primers	CACCCGTTACCCGCATATGA

sequence-based reagent	309_nifD_R	This paper	PCR primers	ACTCATCTGTGAACGGCGTT
sequence-based reagent	310_nifK_F	This paper	PCR primers	GCTAACGCCGTTCACAGATG
sequence-based reagent	311_nifK_R	This paper	PCR primers	TCAGTTGGCCTTCGTCGTTG
software, algorithm	MAFFT	MAFFT	RRID:SCR_011811	
software, algorithm	trimAl	trimAl	RRID:SCR_017334	
software, algorithm	IQ-TREE	IQ-TREE	RRID:SCR_017254	
software, algorithm	RAxML	RAxML	RRID:SCR_006086	
software, algorithm	PAML	PAML	RRID:SCR_014932	

software, algorithm	MODELLER	MODELLER	RRID:SCR_008395	
software, algorithm	ChimeraX	ChimeraX	RRID:SCR_015872	
software, algorithm	Growthcurver	Growthcurver	n/a	R package

342
343
344 **Nitrogenase ancestral sequence reconstruction and selection**
345
346 The nitrogenase protein sequence dataset was assembled by BLASTp [51] search of the NCBI
347 non-redundant protein database (accessed August 2020) with *A. vinelandii* NifH
348 (WP_012698831.1), NifD (WP_012698832.1), and NifK (WP_012698833.1) queries and a 1e-5
349 Expect value threshold (**Supplementary File 1b**). BLASTp hits were manually curated to remove
350 partially sequenced, misannotated, and taxonomically overrepresented homologs. BLASTp hits
351 included protein sequences from homologous Mo-, V-, and Fe-nitrogenase isozymes [11]. H-, D,
352 and K-subunit sequences from these isozymes were individually aligned by MAFFT v7.450 [52]
353 and concatenated along with outgroup dark-operative protochlorophyllide oxidoreductase
354 sequences (Bch/ChlLNB). The final dataset included 385 nitrogenase sequences and 385
355 outgroup sequences. For sequences used to construct Anc1A and Anc2 (internal nodes #960 and
356 #929, respectively), tree reconstruction (using a trimmed alignment generated by trimAl v1.2 [53])
357 and ancestral sequence inference (using the initial untrimmed alignment) were both performed
358 by RAxML v8.2.10 [54] with the LG+G+F evolutionary model (model testing performed by the
359 ModelFinder [55] in the IQ-TREE v.1.6.12 package [56]).
360
361 Due to concerns that RAxML v.8.2 does not implement full, marginal ancestral sequence
362 reconstruction as described by Yang et al. [57], we performed a second phylogenetic analysis as
363 follows. The extant sequence dataset described above was realigned by MAFFT (untrimmed) and

364 tree reconstruction was again performed by RAxML. Ancestral sequence reconstruction was
365 instead performed by PAML v4.9j [58] using the LG+G+F model. From this second reconstruction,
366 Anc1B (internal node #1312), equivalent to Anc1A, was selected for experimental analysis. Anc1B
367 and Anc1A have identical sets of descendent homologs, and their NifD proteins are 95% identical
368 (**Fig. 2*B***).
369
370 Only the ancestral sequences inferred with the most probable residue at each protein site were
371 considered for this study (mean posterior probabilities of targeted nitrogenase subunits range
372 from .95 to .99; see **Supplementary File 1a**). All ancestral sequences were reconstructed from
373 well-supported clades (SH-like aLRT = 99-100 [59]).
374
375 **Ancestral nitrogenase structural modeling and sequence analysis**
376
377 Structural homology models of ancestral sequences were generated by MODELLER v10.2 [60]
378 using PDB 1M34 as a template for all nitrogenase protein subunits and visualized by ChimeraX
379 v1.3 [61].
380
381 Extant and ancestral protein sequence space was visualized by machine-learning embeddings,
382 where each protein embedding represents protein features in a fixed-size, multidimensional
383 vector space. The analysis was conducted on concatenated (HDK) nitrogenase protein
384 sequences in our phylogenetic dataset. The embeddings were obtained using the pre-trained
385 language model ESM2 [62, 63], a transformer architecture trained to reproduce correlations at
386 the sequence level in a dataset containing hundreds of millions of protein sequences. Layer 33 of
387 this transformer was used, as recommended by the authors. The resulting 1,024 dimensions were
388 reduced by UMAP [64] for visualization in a two-dimensional space.
389
390 Protein site-wise conservation analysis was performed using the Consurf server [38]. An input
391 alignment containing only extant, Group I Mo-nitrogenases was submitted for analysis under
392 default parameters. Conserved sites were defined by a Consurf conservation score >7.
393
394 ***A. vinelandii* strain engineering**
395
396 Nucleotide sequences of targeted ancestral nitrogenase proteins were codon-optimized for *A.*
397 *vinelandii* by a semi-randomized strategy that maximized ancestral nucleotide sequence identity

398 to WT genes. Ancestral and WT protein sequences were compared using the alignment output of
399 ancestral sequence reconstruction (RAxML or PAML). For sites where the ancestral and WT
400 residues were identical, the WT codon was assigned. At sites where the residues were different,
401 the codon was assigned randomly, weighted by *A. vinelandii* codon frequencies (Codon Usage
402 Database, https://www.kazusa.or.jp/codon/). Nucleotide sequences were synthesized into
403 XbaI/KpnI sites of pUC19 vectors (unable to replicate in *A. vinelandii*) (Twist Bioscience;
404 GenScript). Inserts were designed with 400-base-pair flanking regions for homology-directed
405 recombination at the relevant *A. vinelandii nif* locus. An "ASWSHPQFEK" Strep-II-tag was
406 included at the N-terminus of each synthetic *nifD* gene for downstream NifD immunodetection
407 and NifDK affinity purification. See **Supplementary File 1c** for a list of strains and plasmids used
408 in this study.
409
410 Engineering of *A. vinelandii* strains used established methods, following Dos Santos [65]. *A.*
411 *vinelandii* WT ("DJ"), DJ2278 (Δ*nifD::*KanR), DJ2102 (Strep-II-tagged WT NifD), and DJ884 (NifH-
412 overexpression mutant) strains were generously provided by Dennis Dean (Virginia Tech)
413 (**Supplementary File 1c**). Strains Anc1A and Anc2 were constructed from the DJ2278 parent
414 strain via transformation with plasmids pAG13 and pAG19, respectively (**Supplementary File**
415 **1c**). For construction of strain Anc2, we first generated a *ΔnifHDK* strain, AK022, by transforming
416 the DJ strain with pAG25. Genetic competency was induced by subculturing relevant parent
417 strains in Mo- and Fe-free Burk's medium (see below) . Competent cells were transformed with
418 at least 1 µg of donor plasmid. Transformants were screened on solid Burk's medium for rescue
419 of the diazotrophic phenotype ("Nif+") and loss of kanamycin resistance, followed by Sanger
420 sequencing of the PCR-amplified *nifHDK* cluster (see **Supplementary File 1d** for a list of
421 primers). Transformants were passaged at least 3 times to ensure phenotypic stability prior to
422 storage at -80 °C in phosphate buffer containing 7% DMSO.
423
424 ***A. vinelandii* culturing and growth analysis**
425
426 *A. vinelandii* strains were grown diazotrophically in nitrogen-free Burk's medium (containing 1 µM
427 Na_2MoO_4) at 30 °C and agitated at 300 rpm. To induce genetic competency for transformation
428 experiments, Mo and Fe salts were excluded. For transformant screening, kanamycin antibiotic
429 was added to solid Burk's medium at a final concentration of 0.6 µg/mL. 50 mL seed cultures for
430 growth rate and acetylene reduction rate quantification were grown non-diazotrophically in flasks
431 with Burk's medium containing 13 mM ammonium acetate.

432

433 For growth rate quantification, seed cultures were inoculated into 100 mL nitrogen-free Burk's
434 medium to an optical density of ~0.01 at 600 nm (OD600), after Carruthers et al. [66], and
435 monitored for 72 h. Growth parameters were modeled using the R package Growthcurver [67].

436

437 **Microbial acetylene reduction assays**

438

439 *A. vinelandii* seed cultures representing independent biological replicates were prepared as
440 described above and used to inoculate 100 mL of nitrogen-free Burk's medium to an OD600 ≈
441 0.01. Cells were grown diazotrophically to an OD600 ≈ 0.5, at which point a rubber septum cap
442 was affixed to the mouth of each flask. 25 mL of headspace was removed and replaced by
443 injecting an equivalent volume of acetylene gas. The cultures were subsequently shaken at 30
444 °C and agitated at 300 rpm. Headspace samples were taken after 15, 30, 45, and 60 mins of
445 incubation for ethylene quantification by a Nexis GC-2030 gas chromatograph (Shimadzu). After
446 the 60 min incubation period, cells were pelleted at 4,700 rpm for 10 mins, washed once with 4
447 mL of phosphate buffer, and pelleted once more under the same conditions prior to storage at -
448 80 °C. Total protein was quantified using the Quick Start™ Bradford Protein Assay kit (Bio-Rad)
449 according to manufacturer instructions and a CLARIOstar Plus plate reader (BMG Labtech).
450 Acetylene reduction rates for each replicate were normalized to total protein.

451

452 **Nitrogenase expression analysis**

453

454 Strep-II-tagged NifD protein quantification was performed on all ancestral strains (Anc1A, Anc1B,
455 Anc2) and DJ2102 (harboring Strep-II-tagged WT NifD). Diazotrophic *A. vinelandii* cultures (100
456 mL) representing three independent biological replicates were prepared as described above and
457 harvested at an OD600 ≈ 1. Cell pellets were resuspended in TE lysis buffer (10 mM Tris, 1 mM
458 EDTA, 1 mg/mL lysozyme) and heated at 95 °C for 10 min. Cell lysates were centrifuged at 5,000
459 rpm for 15 min. Total protein in the resulting supernatant was quantified using the Pierce™ BCA
460 Protein Assay kit (ThermoFisher) following manufacturer instructions. Normalized protein
461 samples were diluted in 2 × Laemmli buffer at a 1:1 (v/v) ratio prior to SDS-PAGE analysis.
462 Proteins were transferred to nitrocellulose membranes (ThermoFisher), stained with Revert 700
463 Total Protein Stain (LI-COR), and imaged on a Odyssey Fc Imager (LI-COR). Membranes were
464 then destained with Revert Destaining Solution (LI-COR) and blocked with 5% non-fat milk in PBS
465 solution (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4.) for 1 h at room

466 temperature. Membranes were rinsed once with PBS-T (PBS with 0.01% Tween-20) and
467 incubated with primary Strep-tag II antibody (Strep-MAB-Classic, IBA Lifesciences, Cat# 2-1507-
468 001, RRID: AB_513133; 1:5000 in 0.2% BSA) for 2 h at room temperature. Membranes were then
469 incubated in LI-COR blocking buffer containing 1:15000 IRDye 680RD Goat anti-Mouse (LI-COR)
470 for 2 h at room temperature and subsequently imaged with an Odyssey Fc Imager (LI-COR).
471 Densitometry analysis was performed with ImageJ [68], with Strep-II-tagged NifD signal intensity
472 normalized to that of the total protein stain.
473
474 **Nitrogenase expression, purification, and biochemical characterization**
475
476 Ancestral nitrogenase NifDK proteins were expressed from relevant *A. vinelandii* strains (Anc1A,
477 Anc1B, Anc2) and purified according to previously published methods [69] with the following
478 modifications: cells were grown diazotrophically in nitrogen-free Burk's medium and no
479 derepression step to a sufficient OD600 (~1.8) before harvesting. WT NifH was expressed in *A.*
480 *vinelandii* strain DJ884 and purified by previously published methods [70]. Protein purity was
481 assessed at ≥ 95% by SDS-PAGE gel with Coomassie blue staining (**Fig. 4-Figure Supplement**
482 **1**).
483
484 Assays were performed in 9.4 mL vials with a MgATP regeneration buffer (6.7 mM MgCl2, 30 mM
485 phosphocreatine, 5 mM ATP, 0.2 mg/mL creatine phosphokinase, 1.2 mg/mL BSA) and 10 mM
486 sodium dithionite in 100 mM MOPS buffer at pH 7.0. Reaction vials were made anaerobic and
487 relevant gases (N_2, C_2H_2, H_2) were added to appropriate concentrations with the headspace
488 balanced by argon. NifDK proteins (~240 kDa) were added to 0.42 μM, the vial vented to
489 atmospheric pressure, and the reaction initiated by addition of NifH (~60 kDa) protein to 8.4 μM.
490 Reactions were run, shaking, at 30˚ C for 8 minutes and stopped by the addition of 500 μL of 400
491 mM EDTA pH 8.0. NH_3 was quantified using a fluorescence protocol [71] with the following
492 modifications: an aliquot of the sample was added to a solution containing 200 mM potassium
493 phosphate pH 7.3, 20 mM o-phthalaldehyde, and 3.5 mM 2-mercaptoethanol and incubated for
494 30 minutes in the dark. Fluorescence was measured at $\lambda_{excitation}$ of 410 nm and $\lambda_{emission}$ of 472 nm
495 and NH_3 was quantified using a standard generated with NH_4Cl. H_2 and C_2H_4 were quantified by
496 gas chromatography by a thermal conductivity detector (GC-TCD) and gas chromatography with
497 flame ionization detector (GC-FID) respectively, according to published methods [72, 73].
498
499

Statistical analyses

Experimental data was statistically analyzed by one-way ANOVA with post-hoc Tukey HSD test.

MATERIALS AVAILABILITY

Physical materials including bacterial strains and plasmids are available to the scientific community upon request.

DATA AND CODE AVAILABILITY

Phylogenetic data, including sequence alignments and phylogenetic trees, and the script for ancestral gene codon-optimization are publicly available at https://github.com/kacarlab/garcia_nif2023. All other data are included as source data and supplementary files.

ACKNOWLEDGEMENTS

We thank Dennis Dean and Valerie Cash for providing *A. vinelandii* strains DJ, DJ2278, DJ2102, and DJ884 and for guidance in genomic manipulations; Jean-Michel Ané, April MacIntyre, and Junko Maeda for guidance and instrumentation support in performing the *in vivo* acetylene reduction assays; Bruno Cuevas for assistance with visualizing nitrogenase sequence space; and the members of the Metal Selection and Utilization Across Eons (MUSE) Consortium for helpful suggestions and discussions. This research was supported by the National Aeronautics and Space Administration (NASA) Interdisciplinary Consortium for Astrobiology Research: Metal Utilization and Selection Across Eons, MUSE (19- ICAR19_2-0007), the University of Wisconsin-Madison College of Agricultural and Life Sciences, the NASA Postdoctoral Program (A.K.G.), NASA Arizona Space Grant (B.M.C.), and the NASA Early Career Faculty Award (B.K.).

COMPETING INTERESTS

The authors declare no competing interests.

534 **SUPPLEMENTARY FILES & LEGENDS**

535

536 garcia_elife_2023_suppfile1.pdf: **Supplementary File 1**

537

538 **Supplementary File 1a. Sequence characteristics of ancestral nitrogenase subunits.**

539

540 **Supplementary File 1b. Host taxa of nitrogenase and outgroup dark-operative**
541 **protochlorophyllide oxidoreductase homologs included for phylogenetic analysis.**

542

543 **Supplementary File 1c. Strains and plasmids used in this study.**

544

545 **Supplementary File 1d. Primers used in this study.**

546

547

548 **TABLES**

549

550 n/a

551

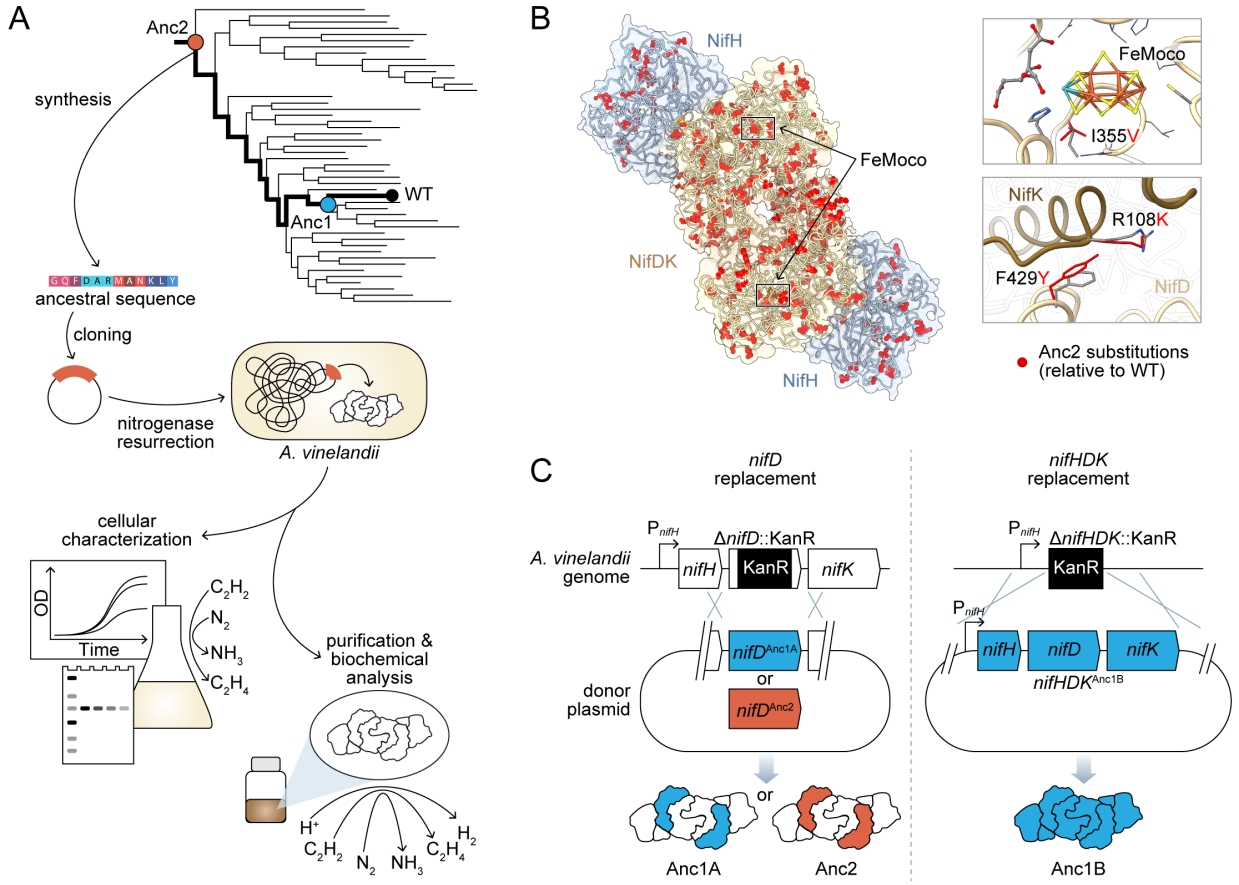
552

553

554 **FIGURES AND FIGURE LEGENDS**

555

556



557

558

559 **Fig. 1. Engineering strategy for ancestral nitrogenase resurrection.** (*A*) Experimental pipeline
560 for nitrogenase resurrection in *A. vinelandii* and subsequent characterization, as described in the
561 main text. (*B*) Structural overview of ancestral nitrogenases reconstructed in this study. Homology
562 models (template PDB 1M34) of Anc1B NifH and NifDK proteins are shown with ancestral
563 substitutions (relative to WT) highlighted in red. Select substitutions at relatively conserved sites
564 in proximity to FeMoco (NifD, I355V) and the NifD:NifK interface (NifD, F429Y; NifK, R108K) are
565 displayed in the insets. (*C*) Parallel genome engineering strategies executed in this study,
566 involving both ancestral replacement of only *nifD* (Anc1A and Anc2) and replacement of *nifHDK*
567 (Anc1B). "P_{nifH}": *nifH* promoter, "KanR": kanamycin resistance cassette. *Anc1A and Anc1B were
568 each reconstructed from equivalent nodes of alternate phylogenies (see **Materials and**
569 **methods**).

570

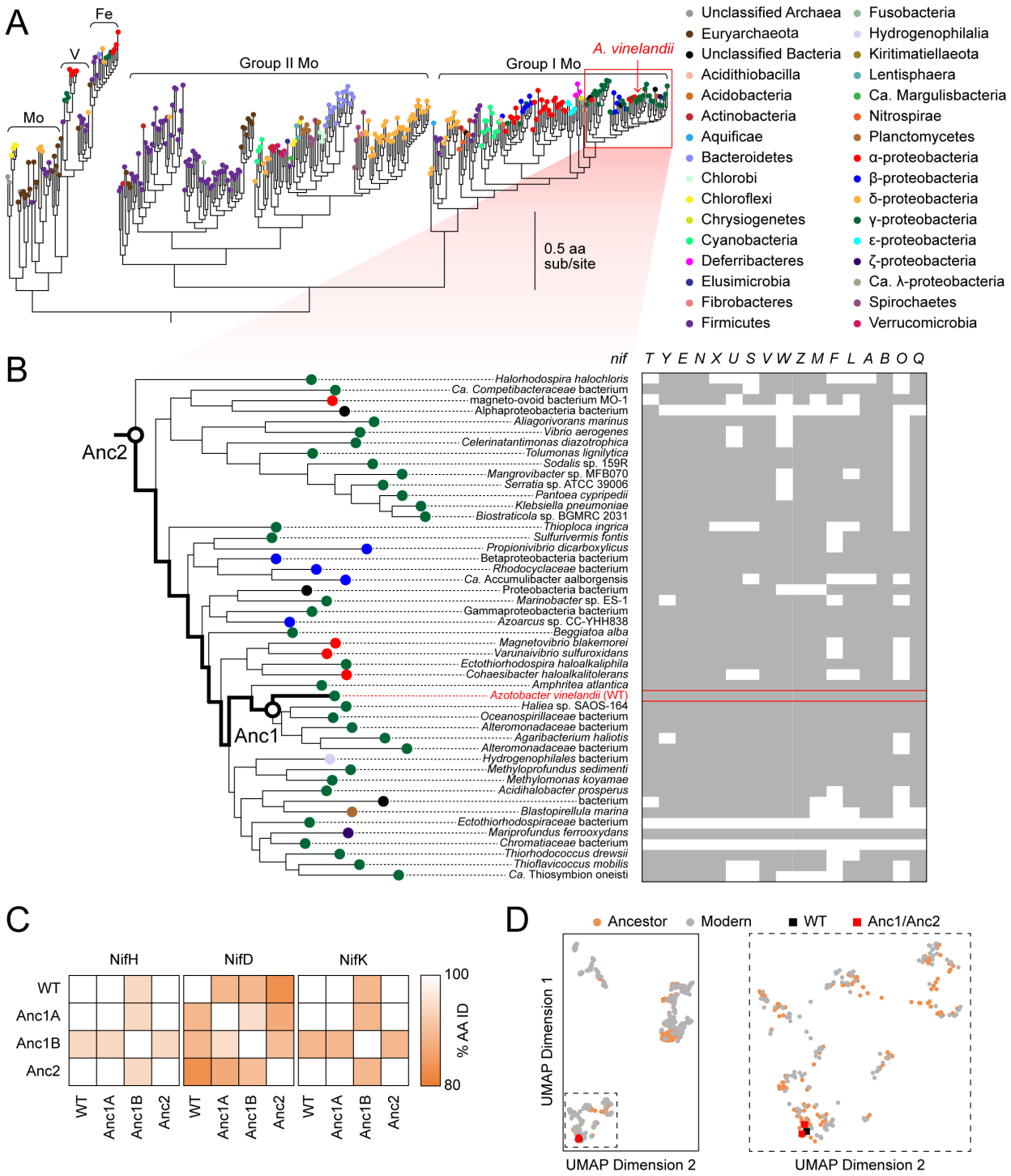


Fig. 2. Phylogenetic and genomic context of resurrected ancestral nitrogenases. (*A*) Maximum-likelihood phylogenetic tree from which ancestral nitrogenases were inferred. Extant nodes are colored by microbial host taxonomic diversity. Red box highlights the clade targeted in this study and depicted in (*B*). Tree shown was used to infer Anc1A and Anc2 sequences (an

577 alternate tree was used for Anc1B inference; see **Materials and methods**). (*B*) *nif* gene cluster
578 complexity within the targeted nitrogenase clade. Presence and absence of each *nif* gene are
579 indicated by grey and white colors, respectively. Because some homologs for phylogenetic
580 analysis were obtained from organisms lacking fully assembled genomes, the absence of
581 accessory *nif* genes may result from missing genomic information. (*C*) Amino acid sequence
582 identity matrix of nitrogenase protein subunits harbored by WT and engineered *A. vinelandii*
583 strains. (*D*) Extant and ancestral nitrogenase protein sequence space visualized by machine-
584 learning embeddings, with the resulting dimensionality reduced to two-dimensional space. UMAP
585 dimension axes are in arbitrary units. The field demarcated by dashed lines in the left plot is
586 expanded on the right plot.
587

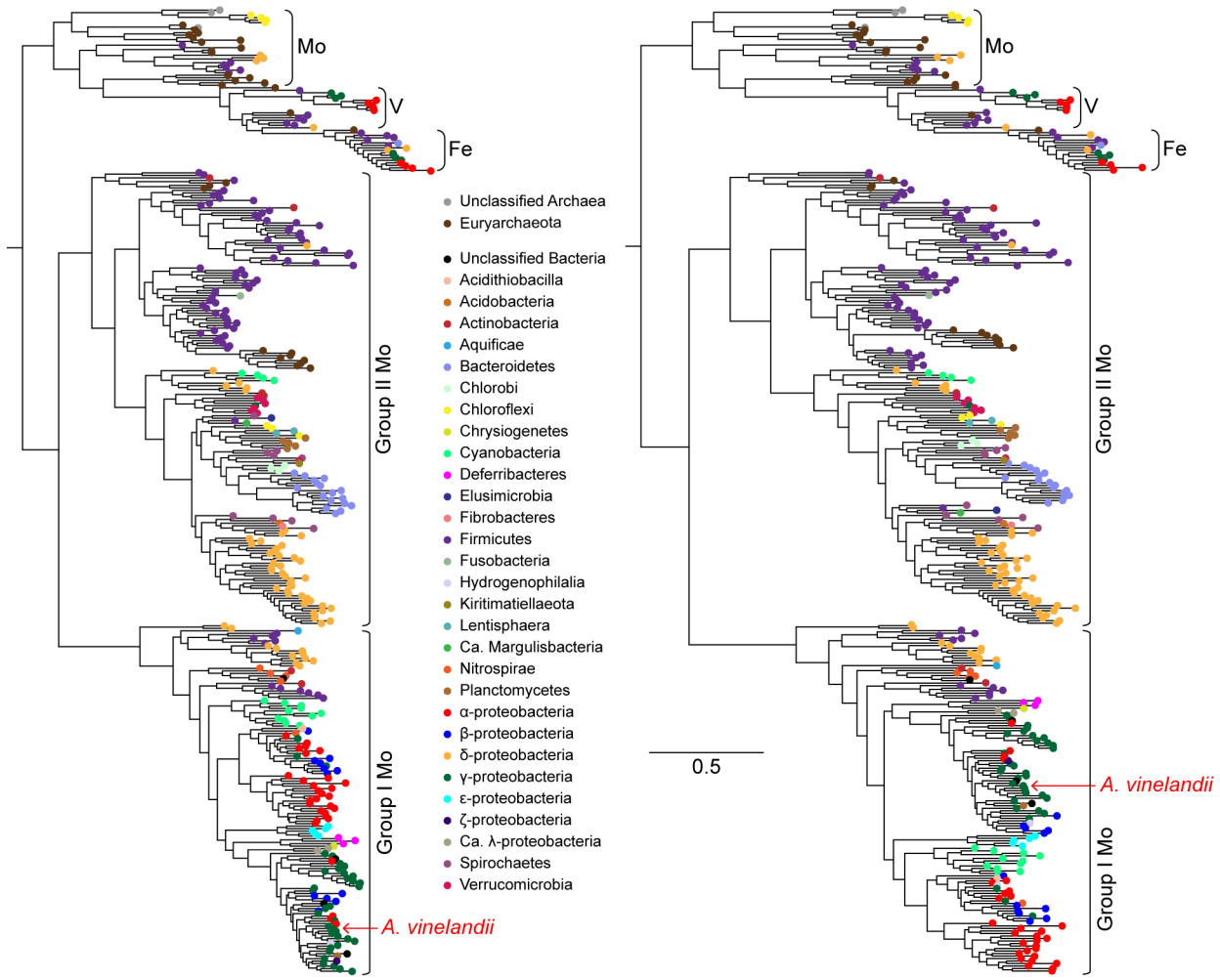


Figure 2-Figure Supplement 1. Maximum-likelihood phylogenies built from nitrogenase NifHDK homologs. Anc1A and Anc2 sequences were inferred from the left tree, and the Anc1B sequence was inferred from the right tree (targeting an equivalent node to Anc1A). Both trees were reconstructed from the same extant sequence dataset (see **Materials and methods** for a description of phylogenetic reconstruction and ancestral sequence inference methods). Branch length scale indicates amino acid substitutions per site and applies to both trees.

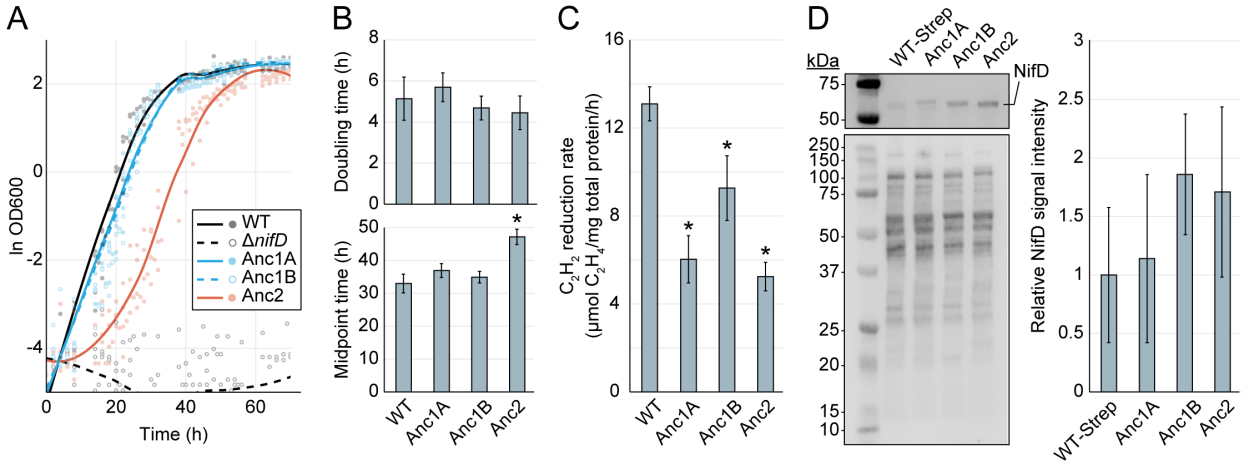


Fig. 3. Cellular-level characterization of ancestral nitrogenase activity and expression. (*A*) Diazotrophic growth curves of *A. vinelandii* strains measured by the optical density at 600 nm ("OD600"). A smoothed curve is shown alongside individual data points obtained from five biological replicates per strain. The non-diazotrophic DJ2278 (Δ*nifD*) strain was used as a negative control. (*B*) Mean doubling and midpoint times of *A. vinelandii* strains, calculated from data in (*A*). (*C*) *In vivo* acetylene (C_2H_2) reduction rates quantified by production of ethylene (C_2H_4). Bars represent the mean of three biological replicates per strain. (*D*) Immunodetection and protein quantification of Strep-II-tagged WT ("WT-Strep", strain DJ2102) and ancestral NifD. Top gel image shows Strep-II-tagged NifD proteins detected by anti-Strep antibody and bottom gel image shows total protein stain. Plot displays relative immunodetected NifD signal intensity normalized to total protein intensity and expressed relative to WT. Bars in plot represent the mean of three biological replicates per strain. (*B-D*) Error bars indicate ± 1 SD and asterisks indicate $p < .01$ (one-way ANOVA, post-hoc Tukey HSD) compared to WT or WT-Strep.

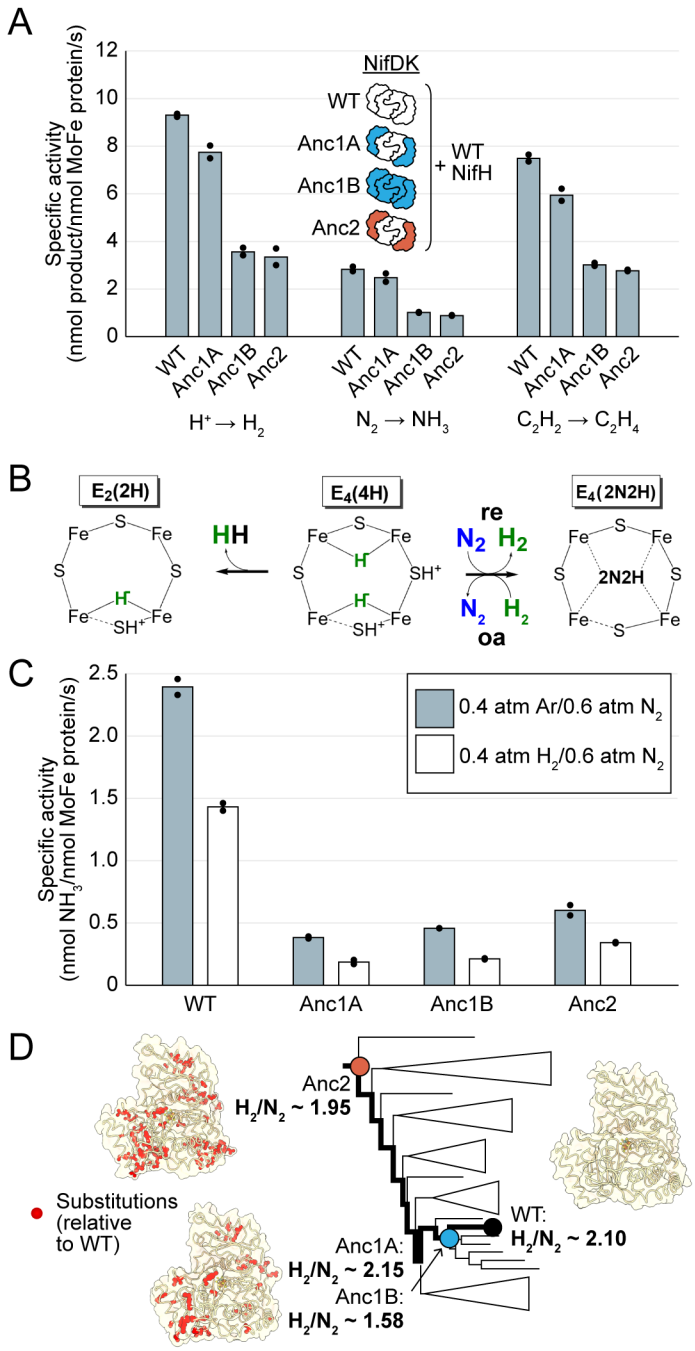


Fig. 4. *In vitro* analyses of ancestral nitrogenase activity profiles and mechanism. All measurements obtained from assays using purified NifDK assayed with WT NifH. (*A*) Specific activities measured for H^+, N_2, and C_2H_2 substrate. (*B*) Partial schematic of the reductive-elimination N_2-reduction mechanism of nitrogenase is shown above, centering on the N_2-binding E_4(4H) state of FeMoco (see main text for discussion) [32]. (*C*) Inhibition of N_2 reduction by H_2, evidencing the mechanism illustrated in (*B*). (*D*) Catalytic efficiencies of ancestral nitrogenases,

629 described by the ratio of formed H_2 to reduced N_2 (H_2/N_2), mapped across the targeted
630 phylogenetic clade. NifD homology models (PDB 1M34 template) are displayed with ancestral
631 substitutions highlighted in red. (*A,C*) Bars represent the mean of two independent experiments
632 with individual data points shown as black circles.
633
634
635

636 **REFERENCES**

637

638 1. Jablonski D. Extinction: past and present. Nature. 2004;427(6975):589. doi:
639 10.1038/427589a.
640 2. Gold DA, Caron A, Fournier GP, Summons RE. Paleoproterozoic sterol biosynthesis and
641 the rise of oxygen. Nature. 2017;543(7645):420-3. Epub 2017/03/07. doi: 10.1038/nature21412.
642 PubMed PMID: 28264195.
643 3. Sanchez-Baracaldo P, Cardona T. On the origin of oxygenic photosynthesis and
644 Cyanobacteria. New Phytol. 2020;225(4):1440-6. Epub 2019/10/11. doi: 10.1111/nph.16249.
645 PubMed PMID: 31598981.
646 4. Goldford JE, Hartman H, Smith TF, Segre D. Remnants of an ancient metabolism without
647 phosphate. Cell. 2017;168(6):1126-34. Epub 2017/03/07. doi: 10.1016/j.cell.2017.02.001.
648 PubMed PMID: 28262353.
649 5. Zerkle AL, House CH, Cox RP, Canfield DE. Metal limitation of cyanobacterial N_2 fixation
650 and implications for the Precambrian nitrogen cycle. Geobiology. 2006;4(4):285-97. doi:
651 10.1111/j.1472-4669.2006.00082.x.
652 6. Soboh B, Boyd ES, Zhao D, Peters JW, Rubio LM. Substrate specificity and evolutionary
653 implications of a NifDK enzyme carrying NifB-co at its active site. FEBS Lett. 2010;584(8):1487-
654 92. Epub 2010/03/12. doi: 10.1016/j.febslet.2010.02.064. PubMed PMID: 20219465.
655 7. Hurley SJ, Wing BA, Jasper CE, Hill NC, Cameron JC. Carbon isotope evidence for the
656 global physiology of Proterozoic cyanobacteria. Sci Adv. 2021;7(2). doi: 10.1126/sciadv.abc8998.
657 8. Garcia AK, Kacar B. How to resurrect ancestral proteins as proxies for ancient
658 biogeochemistry. Free Radic Biol Med. 2019;140:260-9. Epub 2019/04/06. doi:
659 10.1016/j.freeradbiomed.2019.03.033. PubMed PMID: 30951835.
660 9. Kacar B, Guy L, Smith E, Baross J. Resurrecting ancestral genes in bacteria to interpret
661 ancient biosignatures. Philos Trans R Soc A. 2017;375(2109):20160352. Epub 2017/11/15. doi:
662 10.1098/rsta.2016.0352. PubMed PMID: 29133450; PubMed Central PMCID:
663 PMCPMC5686408.
664 10. Kedzior M, Garcia AK, Li M, Taton A, Adam ZR, Young JN, et al. Resurrected Rubisco
665 suggests uniform carbon isotope signatures over geologic time. Cell Rep. 2022;39(4):110726.
666 Epub 2022/04/28. doi: 10.1016/j.celrep.2022.110726. PubMed PMID: 35476992.
667 11. Garcia AK, McShea H, Kolaczkowski B, Kacar B. Reconstructing the evolutionary history
668 of nitrogenases: Evidence for ancestral molybdenum-cofactor utilization. Geobiology.
669 2020;18:394-411. Epub 2020/02/18. doi: 10.1111/gbi.12381. PubMed PMID: 32065506.

670 12. Raymond J, Siefert JL, Staples CR, Blankenship RE. The natural history of nitrogen
671 fixation. Mol Biol Evol. 2004;21(3):541-54. Epub 2003/12/25. doi: 10.1093/molbev/msh047.
672 PubMed PMID: 14694078.
673 13. Boyd ES, Anbar AD, Miller S, Hamilton TL, Lavin M, Peters JW. A late methanogen origin
674 for molybdenum-dependent nitrogenase. Geobiology. 2011;9(3):221-32. Epub 2011/04/21. doi:
675 10.1111/j.1472-4669.2011.00278.x. PubMed PMID: 21504537.
676 14. Stueken EE, Buick R, Guy BM, Koehler MC. Isotopic evidence for biological nitrogen
677 fixation by molybdenum-nitrogenase from 3.2 Gyr. Nature. 2015;520(7549):666-9. Epub
678 2015/02/18. doi: 10.1038/nature14180. PubMed PMID: 25686600.
679 15. Parsons C, Stüeken EE, Rosen CJ, Mateos K, Anderson RE. Radiation of nitrogen-
680 metabolizing enzymes across the tree of life tracks environmental transitions in Earth history.
681 Geobiology. 2020;19:18-34. doi: 10.1111/gbi.12419.
682 16. Garcia AK, Cavanaugh CM, Kacar B. The curious consistency of carbon biosignatures
683 over billions of years of Earth-life coevolution. ISME J. 2021. Epub 2021/04/14. doi:
684 10.1038/s41396-021-00971-5. PubMed PMID: 33846565.
685 17. Mus F, Alleman AB, Pence N, Seefeldt LC, Peters JW. Exploring the alternatives of
686 biological nitrogen fixation. Metallomics. 2018;10(4):523-38. Epub 2018/04/10. doi:
687 10.1039/c8mt00038g. PubMed PMID: 29629463.
688 18. Perez-Gonzalez A, Jimenez-Vicente E, Gies-Elterlein J, Salinero-Lanzarote A, Yang ZY,
689 Einsle O, et al. Specificity of NifEN and VnfEN for the Assembly of Nitrogenase Active Site
690 Cofactors in Azotobacter vinelandii. mBio. 2021;12(4):e0156821. Epub 2021/07/21. doi:
691 10.1128/mBio.01568-21. PubMed PMID: 34281397; PubMed Central PMCID:
692 PMCPMC8406325.
693 19. Buren S, Jimenez-Vicente E, Echavarri-Erasun C, Rubio LM. Biosynthesis of Nitrogenase
694 Cofactors. Chem Rev. 2020;120(12):4921-68. Epub 2020/01/25. doi:
695 10.1021/acs.chemrev.9b00489. PubMed PMID: 31975585; PubMed Central PMCID:
696 PMCPMC7318056.
697 20. Gallon JR. Reconciling the Incompatible - N_2 Fixation and O_2. New Phytol.
698 1992;122(4):571-609. PubMed PMID: WOS:A1992KH15800001.
699 21. Smanski MJ, Bhatia S, Zhao D, Park Y, L BAW, Giannoukos G, et al. Functional
700 optimization of gene clusters by combinatorial design and assembly. Nat Biotechnol.
701 2014;32(12):1241-9. Epub 2014/11/25. doi: 10.1038/nbt.3063. PubMed PMID: 25419741.

702 22. Bennett EM, Murray JW, Isalan M. Engineering Nitrogenases for synthetic nitrogen
703 fixation: From pathway engineering to directed evolution. BioDesign Research. 2023. doi:
704 10.34133/bdr.0005.
705 23. Buren S, Rubio LM. State of the art in eukaryotic nitrogenase engineering. FEMS Microbiol
706 Lett. 2018;365(2). Epub 2017/12/15. doi: 10.1093/femsle/fnx274. PubMed PMID: 29240940;
707 PubMed Central PMCID: PMCPMC5812491.
708 24. Falkowski PG. Evolution of the nitrogen cycle and its influence on the biological
709 sequestration of CO_2 in the ocean. Nature. 1997;387(6630):272-5. doi: 10.1038/387272a0.
710 PubMed PMID: WOS:A1997WZ16700049.
711 25. Allen JF, Thake B, Martin WF. Nitrogenase Inhibition Limited Oxygenation of Earth's
712 Proterozoic Atmosphere. Trends Plant Sci. 2019;24(11):1022-31. Epub 2019/08/27. doi:
713 10.1016/j.tplants.2019.07.007. PubMed PMID: 31447302.
714 26. Vicente EJ, Dean DR. Keeping the nitrogen-fixation dream alive. Proc Natl Acad Sci USA.
715 2017;114(12):3009-11. Epub 2017/03/12. doi: 10.1073/pnas.1701560114. PubMed PMID:
716 28283657; PubMed Central PMCID: PMCPMC5373348.
717 27. Boyd ES, Costas AM, Hamilton TL, Mus F, Peters JW. Evolution of molybdenum
718 nitrogenase during the transition from anaerobic to aerobic metabolism. J Bacteriol.
719 2015;197(9):1690-9. Epub 2015/03/04. doi: 10.1128/JB.02611-14. PubMed PMID: 25733617;
720 PubMed Central PMCID: PMCPMC4403663.
721 28. Mus F, Colman DR, Peters JW, Boyd ES. Geobiological feedbacks, oxygen, and the
722 evolution of nitrogenase. Free Radic Biol Med. 2019. Epub 2019/02/09. doi:
723 10.1016/j.freeradbiomed.2019.01.050. PubMed PMID: 30735835.
724 29. Boyd ES, Hamilton TL, Peters JW. An alternative path for the evolution of biological
725 nitrogen fixation. Front Microbiol. 2011;2:205. Epub 2011/11/09. doi: 10.3389/fmicb.2011.00205.
726 PubMed PMID: 22065963; PubMed Central PMCID: PMCPMC3207485.
727 30. Seefeldt LC, Yang ZY, Lukoyanov DA, Harris DF, Dean DR, Raugei S, et al. Reduction of
728 Substrates by Nitrogenases. Chem Rev. 2020;120:5082-106. Epub 2020/03/17. doi:
729 10.1021/acs.chemrev.9b00556. PubMed PMID: 32176472.
730 31. Stripp ST, Duffus BR, Fourmond V, Leger C, Leimkuhler S, Hirota S, et al. Second and
731 Outer Coordination Sphere Effects in Nitrogenase, Hydrogenase, Formate Dehydrogenase, and
732 CO Dehydrogenase. Chem Rev. 2022;122(14):11900-73. Epub 2022/07/19. doi:
733 10.1021/acs.chemrev.1c00914. PubMed PMID: 35849738.
734 32. Harris DF, Lukoyanov DA, Kallas H, Trncik C, Yang ZY, Compton P, et al. Mo-, V-, and
735 Fe-Nitrogenases Use a Universal Eight-Electron Reductive-Elimination Mechanism To Achieve

736 N_2 Reduction. Biochemistry. 2019;58(30):3293-301. Epub 2019/07/10. doi:
737 10.1021/acs.biochem.9b00468. PubMed PMID: 31283201.
738 33. Noar JD, Bruno-Barcena JM. *Azotobacter vinelandii*: the source of 100 years of
739 discoveries and many more to come. Microbiology. 2018;164(4):421-36. Epub 2018/03/14. doi:
740 10.1099/mic.0.000643. PubMed PMID: 29533747.
741 34. Ghebreamlak SM, Mansoorabadi SO. Divergent Members of the Nitrogenase
742 Superfamily: Tetrapyrrole Biosynthesis and Beyond. ChemBioChem. 2020;21(12):1723-8. Epub
743 2020/03/18. doi: 10.1002/cbic.201900782. PubMed PMID: 32180329.
744 35. Kacar B, Garmendia E, Tuncbag N, Andersson DI, Hughes D. Functional Constraints on
745 Replacing an Essential Gene with Its Ancient and Modern Homologs. MBio. 2017;8(4):e01276-
746 17. Epub 2017/08/31. doi: 10.1128/mBio.01276-17. PubMed PMID: 28851849; PubMed Central
747 PMCID: PMCPMC5574714.
748 36. Lyons TW, Reinhard CT, Planavsky NJ. The rise of oxygen in Earth's early ocean and
749 atmosphere. Nature. 2014;506(7488):307-15. Epub 2014/02/21. doi: 10.1038/nature13068.
750 PubMed PMID: 24553238.
751 37. Dos Santos PC, Dean DR, Hu Y, Ribbe MW. Formation and insertion of the nitrogenase
752 iron-molybdenum cofactor. Chem Rev. 2004;104(2):1159-73. Epub 2004/02/12. doi:
753 10.1021/cr020608l. PubMed PMID: 14871152.
754 38. Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, et al. ConSurf 2016: an
755 improved methodology to estimate and visualize evolutionary conservation in macromolecules.
756 Nucleic Acids Res. 2016;44(W1):W344-50. Epub 2016/05/12. doi: 10.1093/nar/gkw408. PubMed
757 PMID: 27166375; PubMed Central PMCID: PMCPMC4987940.
758 39. Smith BE, Thorneley RN, Eady RR, Mortenson LE. Nitrogenases from *Klebsiella*
759 *pneumoniae* and *Clostridium pasteurianum*. Kinetic investigations of cross-reactions as a probe
760 of the enzyme mechanism. Biochemical Journal. 1976;157(2):439-47. doi: 10.1042/bj1570439.
761 40. Hardy RW, Holsten RD, Jackson EK, Burns RC. The acetylene-ethylene assay for N_2
762 fixation: laboratory and field evaluation. Plant Physiol. 1968;43(8):1185-207. Epub 1968/08/01.
763 doi: 10.1104/pp.43.8.1185. PubMed PMID: 16656902; PubMed Central PMCID:
764 PMCPMC1086994.
765 41. Harris DF, Badalyan A, Seefeldt LC. Mechanistic Insights into Nitrogenase FeMo-Cofactor
766 Catalysis through a Steady-State Kinetic Model. Biochemistry. 2022;61(19):2131-7. Epub
767 2022/09/15. doi: 10.1021/acs.biochem.2c00415. PubMed PMID: 36103672.

768 42. Garcia AK, Kedzior M, Taton A, Li M, Young JN, Kaçar B. System-level effects of CO_2 and
769 RuBisCO concentration on carbon isotope fractionation. bioRxiv, submitted. 2022. doi:
770 10.1101/2021.04.20.440233.
771 43. Som SM, Buick R, Hagadorn JW, Blake TS, Perreault JM, Harnmeijer Jelte P, et al. Earth's
772 air pressure 2.7 billion years ago constrained to less than half of modern levels. Nat Geosci.
773 2016;9(6):448-51. doi: 10.1038/ngeo2713.
774 44. Catling DC, Zahnle KJ. The Archean atmosphere. Sci Adv. 2020;6(9):eaax1420. doi:
775 10.1126/sciadv.aax1420.
776 45. Zehr JP, Jenkins BD, Short SM, Steward GF. Nitrogenase gene diversity and microbial
777 community structure: a cross-system comparison. Environmental Microbiology. 2003;5(7):539-
778 54. doi: 10.1046/j.1462-2920.2003.00451.x.
779 46. Sippel D, Einsle O. The structure of vanadium nitrogenase reveals an unusual bridging
780 ligand. Nat Chem Biol. 2017;13(9):956-60. Epub 2017/07/12. doi: 10.1038/nchembio.2428.
781 PubMed PMID: 28692069; PubMed Central PMCID: PMCPMC5563456.
782 47. Eady RR. Structure-function relationships of alternative nitrogenases. Chemical Reviews.
783 1996;96(7):3013-30. doi: 10.1021/cr950057h.
784 48. Cherkasov N, Ibhadon AO, Fitzpatrick P. A review of the existing and alternative methods
785 for greener nitrogen fixation. Chemical Engineering and Processing: Process Intensification.
786 2015;90:24-33. doi: 10.1016/j.cep.2015.02.004.
787 49. Dorr M, Kassbohrer J, Grunert R, Kreisel G, Brand WA, Werner RA, et al. A possible
788 prebiotic formation of ammonia from dinitrogen on iron sulfide surfaces. Angew Chem Int Ed Engl.
789 2003;42(13):1540-3. Epub 2003/04/17. doi: 10.1002/anie.200250371. PubMed PMID: 12698495.
790 50. Yung YL, McElroy MB. Fixation of nitrogen in the prebiotic atmosphere. Science.
791 1979;203(4384):1002-4. Epub 1979/03/09. doi: 10.1126/science.203.4384.1002. PubMed PMID:
792 17811121.
793 51. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+:
794 architecture and applications. BMC Bioinformatics. 2009;10:421. Epub 2009/12/17. doi:
795 10.1186/1471-2105-10-421. PubMed PMID: 20003500; PubMed Central PMCID:
796 PMCPMC2803857.
797 52. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7:
798 improvements in performance and usability. Mol Biol Evol. 2013;30(4):772-80. Epub 2013/01/19.
799 doi: 10.1093/molbev/mst010. PubMed PMID: 23329690; PubMed Central PMCID:
800 PMCPMC3603318.

801 53. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. trimAl: a tool for automated alignment
802 trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25(15):1972-3. Epub
803 2009/06/10. doi: 10.1093/bioinformatics/btp348. PubMed PMID: 19505945; PubMed Central
804 PMCID: PMCPMC2712344.
805 54. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
806 phylogenies. Bioinformatics. 2014;30(9):1312-3. Epub 2014/01/24. doi:
807 10.1093/bioinformatics/btu033. PubMed PMID: 24451623; PubMed Central PMCID:
808 PMCPMC3998144.
809 55. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast
810 model selection for accurate phylogenetic estimates. Nature Methods. 2017;14(6):587-9. doi:
811 10.1038/nmeth.4285.
812 56. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective
813 stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol.
814 2015;32(1):268-74. Epub 2014/11/06. doi: 10.1093/molbev/msu300. PubMed PMID: 25371430;
815 PubMed Central PMCID: PMCPMC4271533.
816 57. Yang Z, Kumar S, Nei M. A new method of inference of ancestral nucleotide and amino
817 acid sequences. Genetics. 1995;141(4):1641-50. doi: 10.1093/genetics/141.4.1641.
818 58. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol.
819 2007;24(8):1586-91. Epub 2007/05/08. doi: 10.1093/molbev/msm088. PubMed PMID: 17483113.
820 59. Anisimova M, Gascuel O. Approximate likelihood-ratio test for branches: A fast, accurate,
821 and powerful alternative. Syst Biol. 2006;55(4):539-52. Epub 2006/06/21. doi:
822 10.1080/10635150600755453. PubMed PMID: 16785212.
823 60. Webb B, Sali A. Comparative Protein Structure Modeling Using MODELLER. Current
824 Protocols in Bioinformatics. 2016;54(1). doi: 10.1002/cpbi.3.
825 61. Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, et al. UCSF
826 ChimeraX: Structure visualization for researchers, educators, and developers. Protein Sci.
827 2021;30(1):70-82. Epub 2020/09/04. doi: 10.1002/pro.3943. PubMed PMID: 32881101; PubMed
828 Central PMCID: PMCPMC7737788.
829 62. Lin Z, Akin H, Rao R, Hie B, Zhu Z, Lu W, et al. Language models of protein sequences
830 at the scale of evolution enable accurate structure prediction. bioRxiv. 2022. doi:
831 10.1101/2022.07.20.500902.
832 63. Rives A, Meier J, Sercu T, Goyal S, Lin Z, Liu J, et al. Biological structure and function
833 emerge from scaling unsupervised learning to 250 million protein sequences. Proc Natl Acad Sci
834 USA. 2021;118(15). doi: 10.1073/pnas.2016239118.

835 64. McInnes L, Healy J, Melville J. UMAP: Uniform Manifold Approximation and Projection for
836 Dimension Reduction. arXiv. 2020. doi: 10.48550/arXiv.1802.03426.
837 65. Dos Santos PC. Genomic Manipulations of the Diazotroph *Azotobacter vinelandii*.
838 Methods Mol Biol. 2019;1876:91-109. Epub 2018/10/15. doi: 10.1007/978-1-4939-8864-8_6.
839 PubMed PMID: 30317476.
840 66. Carruthers BM, Garcia AK, Rivier A, Kacar B. Automated Laboratory Growth Assessment
841 and Maintenance of *Azotobacter vinelandii*. Curr Protoc Microbiol. 2021;1(3):e57. Epub
842 2021/03/04. doi: 10.1002/cpz1.57. PubMed PMID: 33656286.
843 67. Sprouffske K, Wagner A. Growthcurver: an R package for obtaining interpretable metrics
844 from microbial growth curves. BMC Bioinformatics. 2016;17:172. Epub 2016/04/21. doi:
845 10.1186/s12859-016-1016-7. PubMed PMID: 27094401; PubMed Central PMCID:
846 PMCPMC4837600.
847 68. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image
848 analysis. Nat Methods. 2012;9(7):671-5. Epub 2012/08/30. doi: 10.1038/nmeth.2089. PubMed
849 PMID: 22930834; PubMed Central PMCID: PMCPMC5554542.
850 69. Jimenez-Vicente E, Martin Del Campo JS, Yang ZY, Cash VL, Dean DR, Seefeldt LC.
851 Application of affinity purification methods for analysis of the nitrogenase system from *Azotobacter*
852 *vinelandii*. Methods Enzymol. 2018;613:231-55. Epub 2018/12/05. doi:
853 10.1016/bs.mie.2018.10.007. PubMed PMID: 30509468.
854 70. Christiansen J, Goodwin PJ, Lanzilotta WN, Seefeldt LC, Dean DR. Catalytic and
855 biophysical properties of a nitrogenase Apo-MoFe protein produced by a *nifB*-deletion mutant of
856 *Azotobacter vinelandii*. Biochemistry. 1998;37(36):12611-23. Epub 1998/09/09. doi:
857 10.1021/bi981165b. PubMed PMID: 9730834.
858 71. Corbin JL. Liquid Chromatographic-Fluorescence Determination of Ammonia from
859 Nitrogenase Reactions: A 2-Min Assay. Appl Environ Microbiol. 1984;47(5):1027-30. doi:
860 10.1128/aem.47.5.1027-1030.1984.
861 72. Khadka N, Dean DR, Smith D, Hoffman BM, Raugei S, Seefeldt LC. CO_2 Reduction
862 Catalyzed by Nitrogenase: Pathways to Formate, Carbon Monoxide, and Methane. Inorg Chem.
863 2016;55(17):8321-30. Epub 2016/08/09. doi: 10.1021/acs.inorgchem.6b00388. PubMed PMID:
864 27500789; PubMed Central PMCID: PMCPMC5068488.
865 73. Yang ZY, Dean DR, Seefeldt LC. Molybdenum nitrogenase catalyzes the reduction and
866 coupling of CO to form hydrocarbons. J Biol Chem. 2011;286(22):19417-21. Epub 2011/04/02.
867 doi: 10.1074/jbc.M111.229344. PubMed PMID: 21454640; PubMed Central PMCID:
868 PMCPMC3103320.

869