Nitrogenase resurrection and the evolution of a singular enzymatic mechanism

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

Understanding how life's earliest metabolic innovations emerged and coevolved with the planetary environment is a fundamental challenge for biology and key to unlocking novel bioengineering solutions. Protein resurrection studies can address this challenge through the development of tractable experimental systems to engineer historically consequential metabolisms. These systems must ultimately be tailored to reproduce key features of ancient proteins while permitting their functional evaluation within a complex, cellular and physiochemical context. Here, we present a phylogenetically informed engineering approach to resurrect and characterize ancient nitrogenases within an extant bacterial host. 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. We demonstrate that nitrogenase ancestors both enable diazotrophic growth of the host bacterium and catalyze dinitrogen reduction *in vitro*. Further, we find that the core dinitrogen reduction mechanism conserved by modern nitrogenases was established by the time the studied ancestors emerged. This mechanism is robust to the inferred, historical amino acid substitutions and variable incorporation of one or more ancestral protein subunits. Our results indicate that nitrogenase has been evolutionarily constrained by a single available mechanism for dinitrogen reduction, and has sustained this essential metabolic function 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 (Jablonski, 2004). If life's historical diversity provides a measure of its ability to solve adaptive problems within a coevolving planetary environment, then it too represents an unparalleled repository of past solutions for today's bioengineering challenges. 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 (Gold, Caron, Fournier, & Summons, 2017; Sanchez-Baracaldo & Cardona, 2020; Wolfe & Fournier, 2018), systems-level network reconstructions (Goldford, Hartman, Smith, & Segre, 2017), and the leveraging of extant or mutant biological models as proxies for their ancient counterparts (Hurley, Wing, Jasper, Hill, & Cameron, 2021; Soboh, Boyd, Zhao, Peters, & Rubio, 2010; Thompson et al., 2019; Zerkle, House, Cox, & Canfield, 2006). However, no experimental strategies exist to directly study these metabolic evolutionary histories across past environmental and cellular transitions. 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 (Garcia & Kacar, 2019; Kacar, Guy, Smith, & Baross, 2017; Kedzior et al., 2022). 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, theirbroader metabolic networks, and the overlying 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 (Boyd, Anbar, et al., 2011; Garcia, McShea, Kolaczkowski, & Kacar, 2020; Parsons, Stüeken, Rosen, Mateos, & Anderson, 2020; Raymond, Siefert, Staples, & Blankenship, 2004; Stueken, Buick, Guy, & Koehler, 2015) 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 (Garcia, Cavanaugh, & Kacar, 2021)) is catalyzed by an early evolved (Parsons et al., 2020; Stueken et al., 2015) family of metalloenzymes called nitrogenases that reduce highly inert, atmospheric dinitrogen (N_2) to bioavailable ammonia (NH_3). Though fundamentally reliant on this single enzyme for the N_2 reduction step, many modern prokaryotes also require numerous associated regulatory and biosynthetic genes to sustain nitrogen fixation (Buren, Jimenez-Vicente, Echavarri-Erasun, & Rubio, 2020; Dixon & Kahn, 2004; Inomura, Bragg, & Follows, 2017).

Because nitrogen has been a limiting nutrient over geological timescales (Allen, Thake, & Martin, 2019; Falkowski, 1997; Navarro-Gonzalez, McKay, & Mvondo, 2001; Sanchez-Baracaldo, Ridgwell, & Raven, 2014), 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 (Vicente & Dean, 2017) designed to supplement a remarkable molecular innovation that biology has tinkered with for more than three billion years (Parsons et al., 2020; Stueken et al., 2015). How nitrogen fixation emerged and co-evolved under past environmental conditions is still poorly constrained relative to its importance in Earth's planetary and biological history. For example, how the structural domains and regulatory network of nitrogenase were recruited (Boyd, Costas, Hamilton, Mus, & Peters, 2015; Mus, Colman, Peters, & Boyd, 2019), under what selective pressures the metal dependence of nitrogenases were shaped (Boyd, Hamilton, & Peters, 2011; Garcia et al., 2020), and how the interaction between peptide and metallocluster has been finely tuned to catalyze one of the most difficult reactions in nature (Harris et al., 2019; MacKay & Fryzuk, 2004; Seefeldt et al., 2020; Stripp et al., 2022), all remain open questions. In addition to revealing fundamentally important aspects of metabolic evolution, answering these questions would illuminate the engineering potential of nitrogenase as a whole.

Here, we seek guidance from the Earth's environmental and evolutionary past to explore and understand the evolution of the key metabolic enzyme, nitrogenase. We report an evolutionary systems biology approach for the cellular- and molecular-level characterization of resurrected, ancestral nitrogenases. We describe a proof-of-concept analysis of nitrogenase ancestors engineered 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*. Further biochemical

characterization of purified enzymes shows that ancestors exhibit the core N_2 reduction mechanism retained by their extant counterparts, maintaining the same catalytic selectivity between N_2 and protons. Thus, primary nitrogenase functionality is robust to historical amino acid substitutions reconstructed across the investigated timeline. Our findings with this new model system lay the conceptual and experimental groundwork for further characterization of ancient nitrogen fixation machinery under diverse environmental conditions and across longer timescales.

RESULTS

A resurrection strategy for ancestral nitrogenases

We designed an engineering pipeline for the resurrection and experimental characterization of ancestral nitrogenases (**Fig 1A**). In this scheme, phylogenetically inferred, ancestral nitrogenase genes are synthesized and engineered into the genome of a modern diazotrophic bacterium, enabling the assessment of *in vivo* nitrogenase activity and expression in parallel with biochemical analysis of purified enzyme. The engineering and functional assessment of ancestral nitrogenases required a suitable diazotrophic microbial host, owing to challenges associated with nitrogenase heterologous expression (Kacar, Guy, et al., 2017; Shomar & Bokinsky, 2021; Vicente & Dean, 2017). We selected the obligately aerobic gammaproteobacterium, *A. vinelandii* (strain "DJ"), an ideal experimental model due to its genetic tractability and the availability of detailed studies on the genetics and biochemistry of its nitrogen fixation machinery (Dos Santos, 2019; B. M. Hoffman, Lukoyanov, Yang, Dean, & Seefeldt, 2014; Noar & Bruno-Barcena, 2018). We specifically targeted the extant *A. vinelandii* molybdenum-dependent ("Mo-") nitrogenase (hereafter referred to as wild-type, "WT"), which is the best studied isozyme (Poza-Carrion, Jimenez-Vicente, Navarro-Rodriguez, Echavarri-Erasun, & Rubio, 2014; Seefeldt et al., 2020; Spatzal et al., 2011) relative to the *A. vinelandii* vanadium ("V-") and iron ("Fe-") dependent nitrogenases (Mus, Alleman, Pence, Seefeldt, & Peters, 2018). The WT Mo-nitrogenase complex comprises multiple subunits, NifH, NifD, and NifK (encoded by *nifHDK* genes), which are arranged into two catalytic components: a NifH homodimer and a NifDK heterotetramer (**Fig 1B**). During catalysis, both components transiently associate to transfer one electron from NifH to NifDK and subsequently dissociate. Transferred electrons accumulate at the active-site Mo-containing metallocluster ("FeMoco") housed within the NifD subunits for reduction of N_2 substrate to NH_3.

To infer Mo-nitrogenase ancestors, we built a maximum likelihood nitrogenase phylogeny from a concatenated alignment of NifHDK amino acid sequences (**Fig 2A**). The phylogeny contains 385 sets of homologs representative of known nitrogenase molecular sequence diversity (including Mo-, V-, and Fe-nitrogenases) (Garcia et al., 2020), and is rooted by dark-operative protochlorophyllide oxidoreductase proteins classified within the nitrogenase superfamily (Ghebreamlak & Mansoorabadi, 2020). For our proof-of-concept study, we selected recent ancestors that fall within the direct evolutionary lineage of *A. vinelandii* WT (**Fig 2B**). A relatively conservative percentage identity threshold was chosen based on prior studies benchmarking functional expression of ancestral elongation factor proteins in *Escherichia coli* (Kacar, Garmendia, Tuncbag, Andersson, & Hughes, 2017) and *Synechococcus elongatus* (Kedzior et al., 2022). We selected two ancestral nodes within the WT lineage, “Anc1” and “Anc2” (listed in order of increasing age) having ~90% and ~85% amino acid sequence identity to WT across the full length of their concatenated NifHDK proteins, respectively (**Fig 2C, S1 Table**). The high-dimensional nitrogenase protein sequence space occupied by both extant and ancestral homologs, is visualized in two-dimensions in **Fig 2D** by machine-learning embeddings (see **Materials and methods**). This analysis highlights the swath of sequence space targeted here, as well as that made accessible by resurrection of nitrogenase ancestors more broadly. Age estimates for Anc1 and Anc2 are challenged by rampant horizontal gene transfer within the nitrogenase family (Parsons et al., 2020; Raymond et al., 2004). WT, Anc1, and Anc2 lie within a Mo-nitrogenase clade (previously termed “Group I” (Raymond et al., 2004)) that contain homologs from diverse aerobic and facultatively anaerobic taxa, including proteobacteria and cyanobacteria (**Fig 2A**). A maximum age constraint of ~2.5 Ga for Group I nitrogenases can be reasoned based on timing of the Great Oxidation Event (Lyons, Reinhard, & Planavsky, 2014) and downstream emergence of aerobic taxa represented nearly exclusively within this clade.

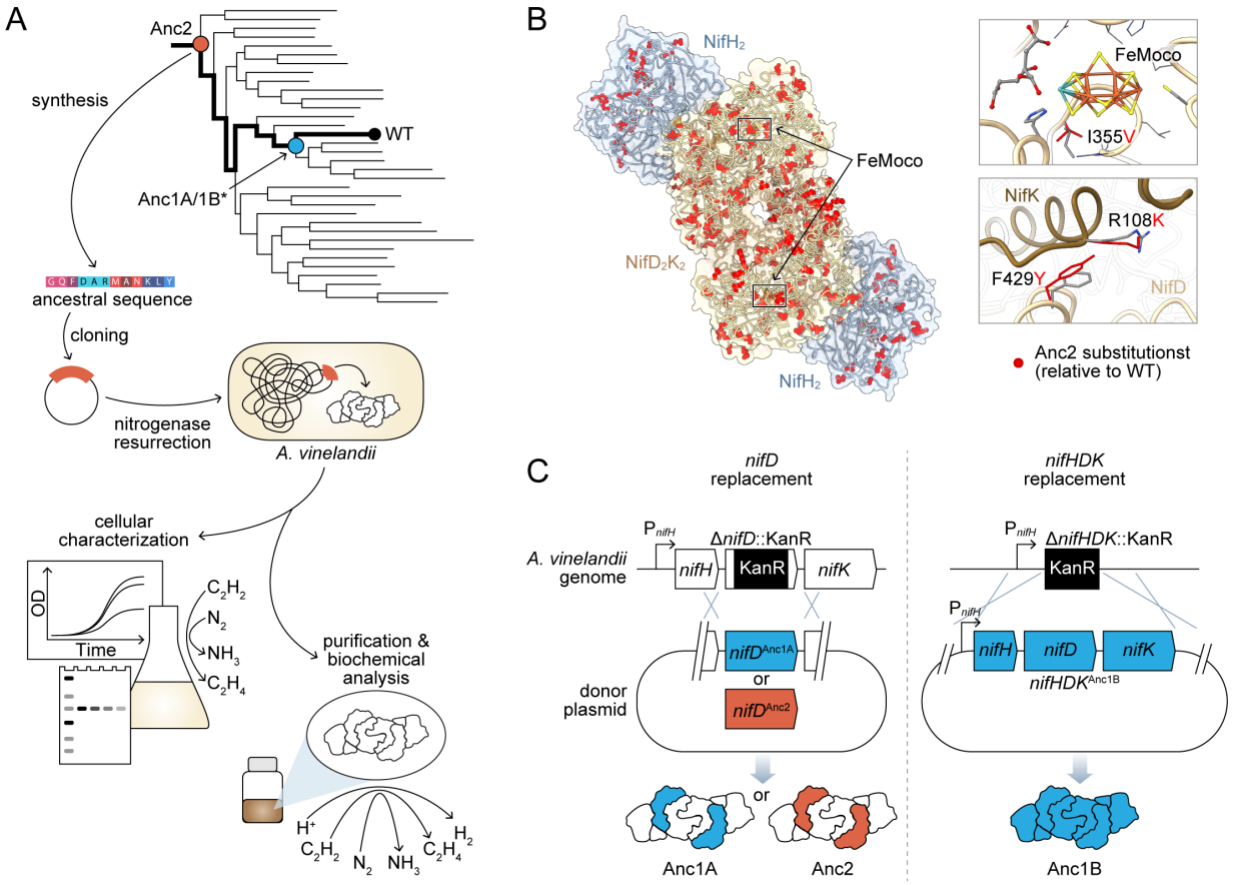

Fig 1. Engineering strategy for ancestral nitrogenase resurrection. (*A*) Experimental pipeline for nitrogenase resurrection in *A. vinelandii* and subsequent characterization, as described in the main text. (*B*) Structural overview of ancestral nitrogenases reconstructed in this study. Homology models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

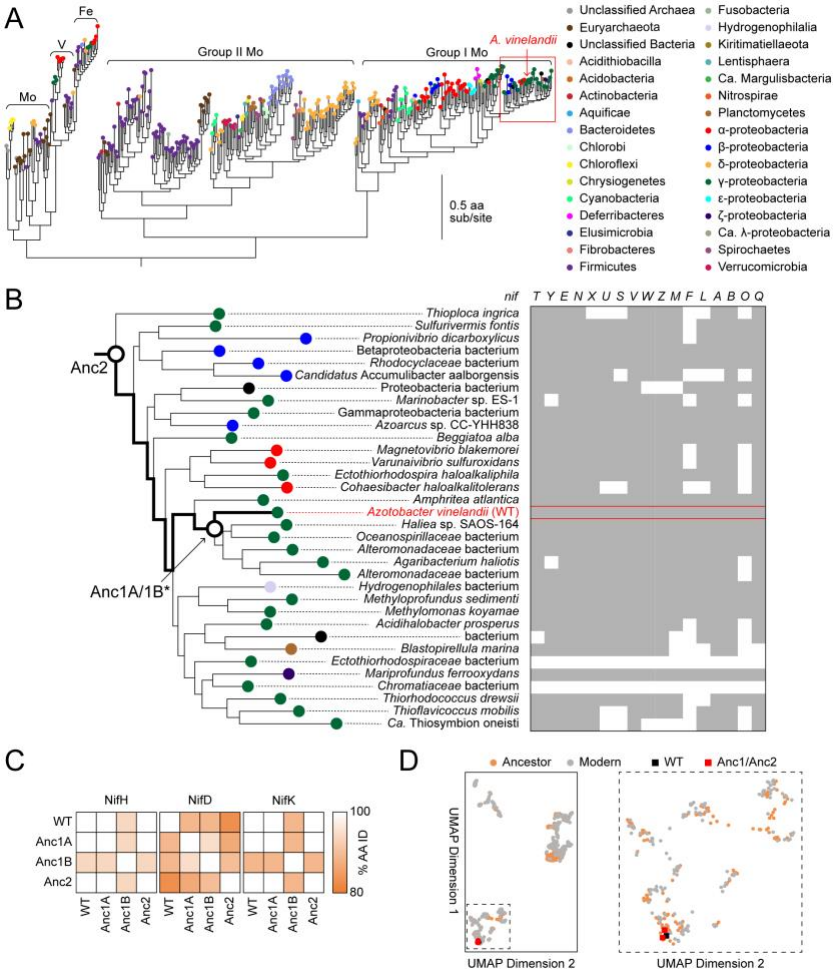

Fig 2. Phylogenetic and genomic context of resurrected ancestral nitrogenases. (*A*) Maximum likelihood phylogenetic tree from which ancestral nitrogenases were inferred. Extant nodes are colored by microbial host taxonomic diversity. Red box highlights the clade targeted in this study and depicted in (*B*). (*B*) *nif* gene cluster complexity within the targeted nitrogenase clade. Presence and absence of each *nif* gene are indicated by grey and white colors, respectively. Because some homologs for phylogenetic analysis were obtained from organisms lacking fully assembled genomes, the absence of accessory *nif* genes may result from missing genomic information. (*C*) Amino acid sequence identity matrix of nitrogenase protein subunits harbored by WT and engineered *A. vinelandii* strains. (*D*) Extant and ancestral nitrogenase protein sequence space visualized by machine-learning embeddings, with the resulting dimensionality reduced to two-dimensional space (see **Materials and methods**). UMAP dimension axes are in arbitrary units. The field demarcated by dashed lines in the left plot is expanded on the right plot. *Anc1A and Anc1B were each reconstructed from equivalent nodes of alternate phylogenies (see **Materials and methods**).

Residue-level differences between ancestral and WT nitrogenases ("ancestral substitutions") are broadly distributed along the length of each ancestral sequence (**Fig 1B, S2 Fig**). The sole ancestral substitution proximal to the active-site FeMoco metallocluster lies within a loop considered important for FeMoco insertion (Dos Santos, Dean, Hu, & Ribbe, 2004) (NifD I355V; residue numbering from WT)) and is observed across all targeted NifD ancestors (**Fig 1B**). Other ancestral substitutions are notable for their location at relatively conserved residue sites (assessed by ConSurf, **S3 Fig**; see **Materials and methods**) and/or within subunit interfaces, including two at the NifD:NifK interface that are proximal to one another, F429Y (NifD) and R108K (NifK). The Anc2 NifD protein contains five more ancestral substitutions at conserved sites than the younger Anc1 NifD protein. In all studied ancestors, the C275 and H442 FeMoco ligands as well as other strictly conserved nitrogenase residues are retained.

Phylogenetic analysis informs the compatibility of selected ancestors in extant microbial hosts. Extant nitrogenases within the Group I nitrogenase clade (which include Anc1 and Anc2 descendants) are associated with numerous accessory genes likely recruited to optimize synthesis and regulation of the oxygen-sensitive nitrogenase for aerobic or facultative metabolisms (Boyd et al., 2015). For example, in addition to the structural *nifHDK* genes, *A. vinelandii* WT is assembled and regulated with the help of >15 *nif* genes (Setubal et al., 2009) (in contrast to the minimal five-*nif*-gene set that can support N-fixation in the anaerobic methanogen *Methanocaldococcus* sp. FS406-22 (Mehta & Baross, 2006)). Likewise, the extant descendants of Anc1 and Anc2 are primarily aerobic or facultative proteobacteria and are thus associated with higher complexity *nif* gene clusters (**Fig 2B**). We hypothesized that the likely oxygen-tolerant, ancient proteobacterium harboring these ancestral nitrogenases were similar in *nif* cluster complexity to extant *A. vinelandii*. Thus, we predicted that the *nif* accessory genes present in *A. vinelandii* would support the functional expression of resurrected nitrogenase ancestors.

To gauge the degree of compatibility between ancestral and extant nitrogenase proteins and maximize the chance of recovering functional nitrogenase variants, we executed two parallel genome engineering strategies. First, we constructed *A. vinelandii* strains harboring only the ancestral *nifD* gene from both targeted ancestral nodes ("Anc1A" and "Anc2"), thereby expressing "hybrid ancestral-WT" nitrogenase complexes (**Fig 1C**). This strategy is similar to *in vitro* "cross-reaction" studies that have evaluated the compatibility of nitrogenase protein components from differing host taxa (Detroy, Witz, Parejko, & Wilson, 1968; B. E. Smith, Thorneley, Eady, & Mortenson, 1976). Second, we constructed a strain harboring all Anc1 *nifHDK* genes, expressing

a fully ancestral nitrogenase complex ("Anc1B"; sequence reconstructed from a node equivalent to Anc1A from an alternate phylogeny, see **Materials and methods**). *A. vinelandii* strains were constructed by markerless genomic integration of ancestral nitrogenase genes, as described in **Materials and methods** (**Table 1**).

Ancestral nitrogenases enable diazotrophic microbial growth

All *A. vinelandii* constructs harboring ancestral genes enabled diazotrophic growth in molybdenum-containing, nitrogen-free media. All strains had comparable doubling times to WT during the exponential phase (**Fig 3A, 3B**). The only significant difference among strains was a ~24-h increase in the lag phase of strain Anc2 relative to WT, harboring the oldest nitrogenase ancestor ($p < 10^{-6}$). We did not detect growth under the same conditions for a control Δ*nifD* strain (DJ2278, see **S2 Table**). This result confirmed that the growth observed for ancestral strains did not stem from leaky expression of the alternative, V- or Fe-dependent nitrogen fixation genes in *A. vinelandii,* which were left intact.

An acetylene reduction assay was performed to measure cellular nitrogenase activity in engineered strains. This assay quantifies the reduction rate of the non-physiological substrate acetylene (C_2H_2) to ethylene (C_2H_4) (Hardy, Holsten, Jackson, & Burns, 1968), here normalized to total protein content. *A. vinelandii* strains harboring only ancestral *nifD* (Anc1A, Anc2) exhibited mean C_2H_2 reduction rates of ~5 to 6 μmol C_2H_4/mg total protein/h, ~50% that of WT ($p < .01$) (**Fig 3C**). Strain Anc1B, harboring ancestral *nifHDK,* had comparable acetylene reduction rates to WT.

The phenotypic variability we observed among engineered and WT *A. vinelandii* strains might result both from differences in nitrogenase expression and nitrogenase activity. To provide insights into these disparate effects, we quantified nitrogenase protein expression in engineered strains by immunodetection of ancestral and WT Strep-tagged NifD proteins (the latter from strain DJ2102, see **S2 Table**) and did not conclusively detect significant differences in protein quantity relative to WT (**Fig 3D**).

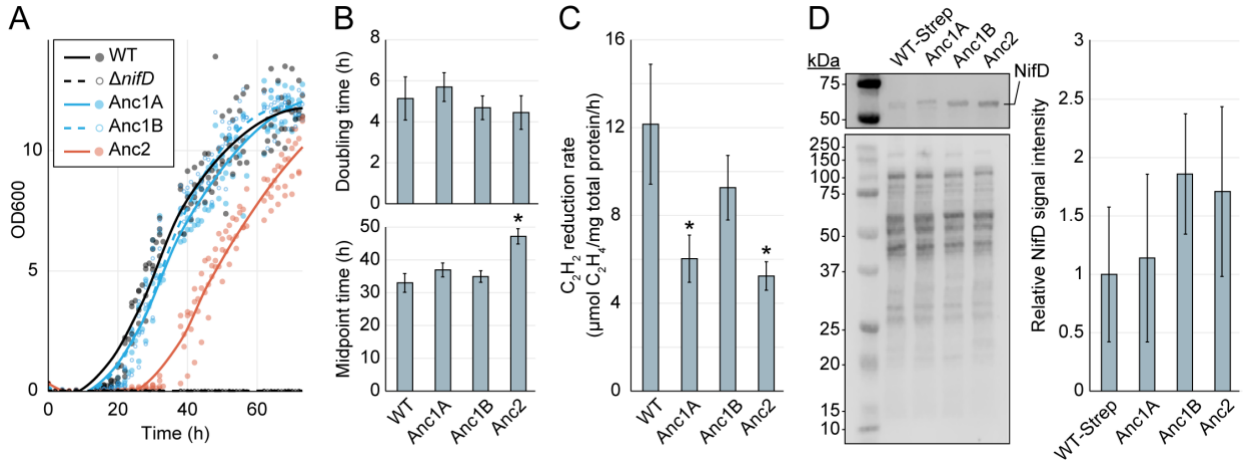

Fig 3. Cellular-level characterization of ancestral nitrogenase activity and expression. (*A*) Diazotrophic growth curves of *A. vinelandii* strains measured by the optical density at 600 nm ("OD600"). A smoothed curve is shown alongside individual data points obtained from five biological replicates 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") and ancestral NifD (Strep-tagged WT NifD strain, DJ2102, used as control). 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 $p < .01$ compared to WT or WT-Strep.

Purified ancestral nitrogenases conserve extant N_2 reduction mechanism and efficiency

Ancestral nitrogenase NifDK protein components were expressed and purified for biochemical characterization. All ancestral NifDK proteins were assayed with WT NifH protein (ancestral NifH proteins were not purified) for reduction of H^+, N_2, and C_2H_2. Ancestors were found to reduce all three substrates *in vitro*, supporting the cellular-level evidence of ancestral nitrogenase activity (**Fig 4A**).

We investigated whether the ancestral nitrogenases studied here would conserve the general mechanism for N_2 binding and reduction that has been observed for all studied, extant nitrogenase isozymes (Mo, V, and Fe) (Harris et al., 2019; Seefeldt et al., 2020). This mechanism involves the accumulation of 4 electron/protons on the active-site cofactor as metal-bound hydrides, generating the E_4(4H) state (**Fig 4B**). Once generated, N_2 can bind to the E_4(4H) state through a reversible reductive elimination/oxidative addition (*re*/*oa*) mechanism, which results in the release (*re*) of a single molecule of hydrogen gas (H_2). N_2 binding is reversible in the presence of sufficient H_2, which displaces bound N_2 and results in reformation of E_4(4H) with two hydrides (*oa*). Thus, a classic test of the (*re*/*oa*) mechanism is the ability of H_2 to inhibit N_2 reduction. We observed that the reduction of N_2 to NH_3 for all nitrogenase ancestors was inhibited in the presence of H_2, indicating that the ancestors follow the same mechanism of N_2 binding determined for extant enzymes (**Fig 4C**).

In the event the E_4(4H) state fails to capture N_2, nitrogenases will simply produce H_2 from the 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 used as a measure of the efficiency of nitrogenases in using ATP and reducing equivalents for N_2 reduction. The stoichiometric minimum of the mechanism is $H_2/N_2 = 1$. Experimentally (under 1 atm N_2), a ratio of ~2 is seen for Mo-nitrogenase and ~5 and ~7 for V- and Fe-nitrogenase, respectively (Harris et al., 2019). H_2/N_2 values for all ancestors under 1 atm N_2 was ~2, similar to extant Mo-nitrogenase (**Fig 4D**).

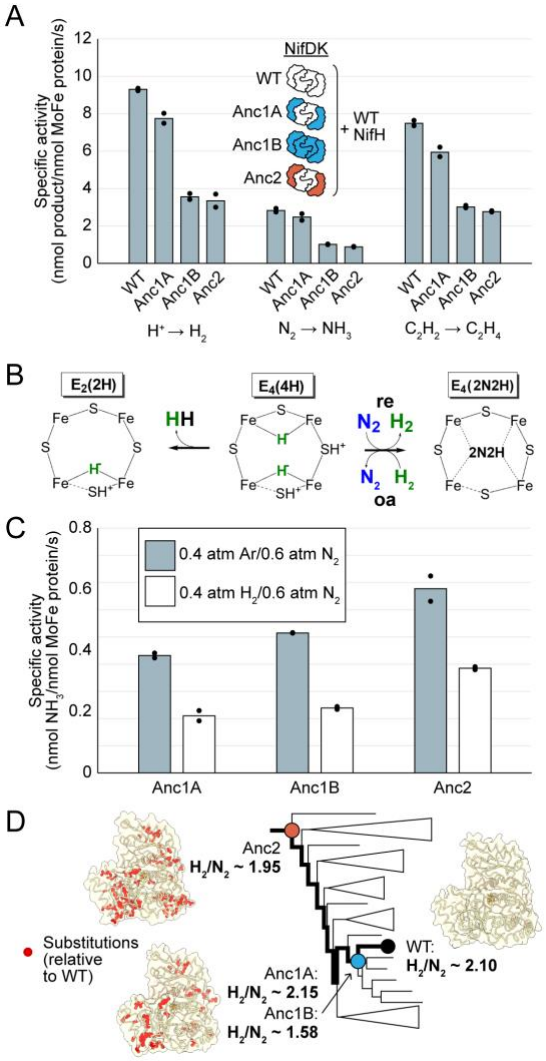

Fig 4. *In vitro* analyses of ancestral nitrogenase activity profiles and mechanism. All measurements obtained from assays using purified NifDK assayed with WT NifH. (*A*) Specific activities measured for H^+, N_2, and C_2H_2 substrate. Bars represent the mean of two independent experiments. (*B*) Partial 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) (Harris et al., 2019). (*C*) Inhibition of N_2 reduction by H_2, evidencing the mechanism illustrated in (*B*). (*D*) Catalytic efficiencies of ancestral nitrogenases, described by the ratio of formed H_2 to reduced N_2 (H_2/N_2), mapped across the targeted phylogenetic clade. NifD homology models (PDB 1M34 template) are displayed with ancestral substitutions highlighted in red. (*A*-*B*) Bars represent the mean of two independent experiments with individual data points shown in black circles. Error bars indicate ± 1 SD.

DISCUSSION

In this study, we describe a new approach to investigate ancient nitrogen fixation by the resurrection and functional assessment of ancestral nitrogenase enzymes. We demonstrate that engineered *A. vinelandii* cells can reduce N_2 and C_2H_2 and exhibit diazotrophic growth rates comparable to WT, though we observe that the oldest ancestor, Anc2, has a significantly longer lag phase. Purified nitrogenase ancestors are active for reduction of H^+, N_2, and C_2H_2, while maintaining the catalytic efficiency (described by the H_2/N_2 ratio) of WT enzymes. Our results also indicate that ancestral N_2 reduction is inhibited by H_2, indicating an early emergence of the reductive-elimination N_2-reduction mechanism preserved by extant nitrogenases (Harris et al., 2019). Intriguingly, these properties are maintained despite substantial residue-level changes to the peripheral nitrogenase structure, as well as a handful within relatively conserved, active-site or protein-interface regions within the enzyme complex.

The observed N_2-reduction activity of nitrogenase ancestors suggests that the required protein-protein interactions—both between subunits that comprise the nitrogenase complex as well as those required for nitrogenase assembly in *A. vinelandii*—and metallocluster interactions are sufficiently maintained for primary function. A striking feature of our results is the phenotypic similarity between strains harboring single (Anc1A) versus multiple (Anc1B) ancestral nitrogenase subunits of equivalent age and genetic distance to WT. The additional incorporation of ancestral NifH and NifK in Anc1B appears to yield comparable diazotrophic growth rates and *in vivo* acetylene reduction rates to Anc1A, whereas both capabilities are diminished in the Anc2 strain hosting a single, older ancestral NifD variant. If borne out by future studies conducted across longer evolutionary transects, these results may suggest that nitrogenase function is more permissive of residue-level perturbations within NifH and NifK than NifD. A deeper analysis is required to identify the specific protein features that impact relative sequence constraints between individual nitrogenase subunits.

That nitrogenase ancestors perform the reductive-elimination N_2-reduction mechanism, as distantly related (Garcia et al., 2020), extant Mo-, V-, and Fe-nitrogenases do today (Harris et al., 2019), likely indicates that this enzymatic characteristic was set early in nitrogenase evolutionary history. If confirmed to permeate nitrogenase ancestry through the approach laid here, it would follow that that this mechanistic scheme has been sustained over billions of years, through significant past environmental change (Catling & Zahnle, 2020; P. F. Hoffman, Kaufman,

Halverson, & Schrag, 1998; Lyons et al., 2014; Som et al., 2016) and ecological diversification (Boyd et al., 2015; Dos Santos, Fang, Mason, Setubal, & Dixon, 2012; Zehr, Jenkins, Short, & Steward, 2003). 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 (Garcia et al., 2020), nitrogenases may have required substantial sequence and structural changes (Eady, 1996; Harris et al., 2018; Sippel & Einsle, 2017) in order to facilitate reductive-elimination given a different active-site metallocluster.

The experimental, paleomolecular approach we have presented here provides a platform to test such hypotheses regarding the evolutionary landscape of nitrogen fixation machinery. 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 (Seefeldt et al., 2020; Stripp et al., 2022), though in certain cases improved reactivity toward alternate, industrially relevant substrates (Fixen et al., 2016; Seefeldt et al., 2020). 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. The use of phylogenetic information as a guide for experimental investigations provides a way to efficiently traverse functional sequence space already "vetted" by natural selection (**Fig 2D**). These strategies can integrate the effects of epistatic interactions that are challenging to predict *a priori* and are thus more likely to produce functional variants than rationally designed or random strategies that can be hindered by local adaptive valleys (Castle, Grierson, & Gorochowski, 2021; Clifton, Kozome, & Laurino, 2022; Hochberg & Thornton, 2017; Spence, Kaczmarski, Saunders, & Jackson, 2021).

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. Enzyme resurrection has provided a means to directly test paradigms of early protein and metabolic evolution, including those associated with the emergence, robustness, and malleability of protein functionality (Copley, 2021; Ferrada & Wagner, 2008; Harms et al., 2013; Ohno, 1970; Risso, Gavira, & Sanchez-Ruiz, 2014), the development of protein-protein interactions and complexes (Eck & Dayhoff, 1966; Jemth et al., 2018; Schulz, Sendker, & Hochberg, 2022), and the role of evolutionary contingency (Blount,

Lenski, & Losos, 2018; Kacar, Ge, Sanyal, & Gaucher, 2017; Venkataram, Monasky, Sikaroodi, Kryazhimskiy, & Kacar, 2020). Nevertheless, 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 (Garcia & Kacar, 2019; Kedzior et al., 2022). Importantly, links between molecular and geological records of life rely on organism- or population-level outputs (e.g., organic biomarkers and isotopic compositions of preserved organic matter) (Garcia et al., 2021; Love et al., 2009; Stüeken, Kipp, Koehler, & Buick, 2016). Further, life's toolbox for adapting nitrogen fixation capabilities under different selective pressures has included the expansion of nitrogenase-associated biosynthetic and regulatory pathways (Boyd et al., 2015; Gallon, 1992) alongside modifications to the sequence-structure features of the nitrogenase enzyme itself. Resurrecting different components of nitrogen fixation machinery in an integrated experimental system can yield insights into the evolutionary history of the machinery as a whole, as well as identify tractable genetic targets for bioengineering.

Our results establish the tractability of leveraging phylogenetic models to carry out extensive, empirical manipulations of challenging enzymatic systems and their microbial hosts. The effects of evolutionary constraints on the development of metabolic pathways that shape the biosphere remain underexplored. Building on empirical frameworks such as that presented here can be used to both reveal the evolutionary constraints of ancient metabolic systems as well as leverage these histories to inform future biological innovations.

MATERIALS AND METHODS

Nitrogenase ancestral sequence reconstruction and selection

The nitrogenase protein sequence dataset was assembled by BLASTp (Camacho et al., 2009) search of the NCBI non-redundant protein database (accessed August 2020) with a 1e-5 Expect value threshold (**S2 Table**). BLASTp hits were manually curated to remove partially sequenced, misannotated, and taxonomically overrepresented homologs. NifH, NifD, and NifK subunit sequences were individually aligned and concatenated along with outgroup dark-operative protochlorophyllide oxidoreductase sequences (Bch/ChlLNB), producing a final 770-sequence dataset. For sequences used to construct Anc1A and Anc2, tree reconstruction and marginal ancestral sequence inference were performed consecutively by RAxML v8.2.10 (Stamatakis,

2014) using the LG+G+F evolutionary model. For Anc1B (inferred from an equivalent node to Anc1A), tree inference was repeated in RAxML using the same extant sequence dataset and model parameters as described above, with ancestral sequence reconstruction performed by PAML v4.9j (Z. Yang, 2007), also using the LG+G+F model. Thus, two alternative nitrogenase phylogenies were incorporated in this study. The nodes used to reconstruct Anc1B and Anc1A have identical sets of descendent homologs, and the sequences themselves are 95% identical across their NifHDK amino acid sequences (**Fig 2B**). Only the ancestral sequences inferred with the most probable residue at each protein site were considered for this study (mean posterior probabilities of targeted nitrogenase subunits range from .95 to .99; see **S1 Table**).

Ancestral nitrogenase structural modeling and sequence analysis

Structural homology models of ancestral sequences were generated by MODELLER v10.2 (Webb & Sali, 2016) using PDB 1M34 as a template for all nitrogenase protein subunits and visualized by ChimeraX v1.3 (Pettersen et al., 2021).

Extant and ancestral protein sequence space was visualized by machine-learning embeddings, where each protein embedding represents protein features in a fixed-size, multidimensional vector space. The embeddings were obtained using the pre-trained language model ESM2 (Lin et al., 2022; Rives et al., 2021), a transformer architecture trained to reproduce correlations at the sequence level in a dataset containing hundreds of millions of protein sequences. Layer 33 of this transformer was used, as recommended by the authors. The resulting 1,024 dimensions were reduced by UMAP (McInnes, Healy, & Melville, 2020) for visualization in a two-dimensional space.

Protein sitewise conservation analysis was performed using the Consurf server (Ashkenazy et al., 2016). An input alignment containing only extant, Group I Mo-nitrogenases was submitted for analysis under default parameters. Conserved sites were defined by a Consurf conservation score >7.

***A. vinelandii* strain engineering**

Nucleotide sequences of targeted ancestral nitrogenase proteins were codon-optimized for *A. vinelandii*. To maximize ancestral nucleotide sequence identity to WT genes, randomly optimized synonymous codons were subsequently replaced by their WT counterparts. Nucleotide

sequences were synthesized into XbaI/KpnI sites of pUC19 vectors (unable to replicate in *A. vinelandii*) (Twist Bioscience; GenScript). Inserts were designed with 400 base-pair flanking regions for homology-directed recombination at the relevant *A. vinelandii nif* locus. An "ASWSHPQFEK" Strep-II-tag was included at the N-terminus of each synthetic *nifD* gene for downstream NifD immunodetection and NifDK affinity purification. See **S3 Table** for a list of strains and plasmids used in this study.

Engineering of *A. vinelandii* strains used established methods, following Dos Santos (Dos Santos, 2019). *A. vinelandii* WT ("DJ"), DJ2278 (Δ*nifD::*KanR), DJ2102 (Strep-II-tagged WT NifD), and DJ884 (NifH-overexpression mutant) strains were generously provided by Dennis Dean (Virginia Tech) (**S3 Table**). Strains Anc1A and Anc2 were constructed from the DJ2278 parent strain via transformation with plasmids pAG13 and pAG19, respectively (**S3 Table**). For construction of strain Anc2, we first generated a *ΔnifHDK* strain, AK022, by transforming the DJ strain with pAG25. Genetic competency was induced by subculturing relevant parent strains in Mo- and Fe-free Burk's medium (see below) (Page & von Tigerstrom, 1979). Competent cells were transformed with at least 1 µg of donor plasmid. Transformants were screened on solid Burk's medium for rescue of the diazotrophic phenotype ("Nif+") and loss of kanamycin resistance, followed by Sanger sequencing of the PCR-amplified *nifHDK* cluster (see **S4 Table** for a list of primers). Transformants were passaged at least 3 times to ensure phenotypic stability prior to storage at -80 °C in phosphate buffer containing 7% DMSO.

A. vinelandii **culturing and growth analysis**

A. vinelandii strains were grown diazotrophically in nitrogen-free Burk's medium (containing 1 µM Na_2MoO_4) at 30 °C and agitated at 300 rpm. To induce genetic competency for transformation experiments, Mo and Fe salts were excluded. For transformant screening, kanamycin antibiotic was added to solid Burk's medium at a final concentration of 0.6 µg/mL. 50 mL seed cultures for growth rate and acetylene reduction rate quantification were grown non-diazotrophically in flasks with Burk's medium containing 13 mM ammonium acetate.

For growth rate quantification, seed cultures were inoculated into 100 mL nitrogen-free Burk's medium to an optical density of ~0.01 at 600 nm (OD600), after Carruthers et al. (Carruthers, Garcia, Rivier, & Kacar, 2021), and monitored for 72 h. Growth parameters were modeled using the R package Growthcurver (Sprouffske & Wagner, 2016).

Microbial acetylene reduction assays

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

Nitrogenase expression analysis

Strep-II-tagged NifD protein quantification was performed on all ancestral strains (Anc1A, Anc1B, Anc2) and DJ2102 (harboring Strep-II-tagged WT NifD). Diazotrophic *A. vinelandii* cultures (100 mL) representing three independent biological replicates were prepared as described above and harvested at an OD600 ≈ 1. Cell pellets were resuspended in TE lysis buffer (10 mM Tris, 1 mM EDTA, 1 mg/mL lysozyme) and heated at 95 °C for 10 min. Cell lysates were centrifuged at 5,000 rpm for 15 min. Total protein in the resulting supernatant was quantified using the Pierce™ BCA Protein Assay kit (ThermoFisher) following manufacturer instructions. Normalized protein samples were diluted in 2 × Laemmli buffer at a 1:1 (v/v) ratio prior to SDS-PAGE analysis. Proteins were transferred to nitrocellulose membranes (ThermoFisher), stained with Revert 700 Total Protein Stain (LI-COR), and imaged on a Odyssey Fc Imager (LI-COR). Membranes were then destained with Revert Destaining Solution (LI-COR) and blocked with 5% non-fat milk in PBS solution (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4.) for 1 h at room temperature. Membranes were rinsed once with PBS-T (PBS with 0.01% Tween-20) and incubated with primary Strep-tag II antibody (Strep-MAB, IBA Lifesciences, 1:5000 in 0.2% BSA) for 2 h at room temperature. Membranes were then incubated in LI-COR blocking buffer containing 1:15000 IRDye 680RD Goat anti-Mouse (LI-COR) for 2 h at room temperature and

subsequently imaged with an Odyssey Fc Imager (LI-COR). Densitometry analysis was performed with ImageJ (Schneider, Rasband, & Eliceiri, 2012), with Strep-II-tagged NifD signal intensity normalized to that of the total protein stain.

Nitrogenase expression, purification, and biochemical characterization

Ancestral nitrogenase NifDK proteins were expressed from relevant *A. vinelandii* strains (Anc1A, Anc1B, Anc2) and purified according to previously published methods (Jimenez-Vicente et al., 2018) with the following modifications: cells were grown diazotrophically in nitrogen-free Burk's medium and no derepression step to a sufficient OD600 (~1.8) before harvesting. WT NifH was expressed in *A. vinelandii* strain DJ884 and purified by previously published methods (Christiansen, Goodwin, Lanzilotta, Seefeldt, & Dean, 1998). Protein purity was assessed at ≥ 95% by SDS-PAGE gel with Coomassie blue staining (**S4 Fig**).

Assays were performed in 9.4 mL vials with a MgATP regeneration buffer (6.7 mM MgCl2, 30 mM phosphocreatine, 5 mM ATP, 0.2 mg/mL creatine phosphokinase, 1.2 mg/mL BSA) and 10 mM sodium dithionite in 100 mM MOPS buffer at pH 7.0. Reaction vials were made anaerobic and relevant gases (N_2, C_2H_2, H_2) were added to appropriate concentrations with the headspace balanced by argon. NifDK proteins (~240 kDa) were added to 0.42 µM, the vial vented to atmospheric pressure, and the reaction initiated by addition of NifH (~60 kDa) protein to 8.4 µM. Reactions were run, shaking, at 30˚ C for 8 minutes and stopped by the addition of 500 µL of 400 mM EDTA pH 8.0. NH_3 was quantified using a fluorescence protocol (Corbin, 1984) with the following modifications: an aliquot of the sample was added to a solution containing 200 mM potassium phosphate pH 7.3, 20 mM o-phthalaldehyde, and 3.5 mM 2-mercaptoethanol and incubated for 30 minutes in the dark. Fluorescence was measured at $\lambda_{excitation}$ of 410 nm and $\lambda_{emission}$ of 472 nm and NH_3 was quantified using a standard generated with NH_4Cl. H_2 and C_2H_4 were quantified by gas chromatography by a thermal conductivity detector (GC-TCD) and gas chromatography with flame ionization detector (GC-FID) respectively, according to published methods (Khadka et al., 2016; Z. Y. Yang, Dean, & Seefeldt, 2011).

Statistical analyses

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

DATA AVAILABILITY

Phylogenetic data, including sequence alignments and phylogenetic trees, are publicly available at https://github.com/kacarlab/garcia_nif2022.

ACKNOWLEDGEMENTS

We thank Dennis Dean and Valerie Cash for providing *A. vinelandii* strains DJ, DJ2278, DJ2102, and DJ884 and for guidance in genomic manipulations, as well as 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 MUSE Consortium for helpful suggestions and discussions. This research was supported by the NASA Interdisciplinary Consortium for Astrobiology Research: Metal Utilization and Selection Across Eons, MUSE (19- ICAR19_2-0007), the UW-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.

REFERENCES

Allen, J. F., Thake, B., & Martin, W. F. (2019). Nitrogenase Inhibition Limited Oxygenation of Earth's Proterozoic Atmosphere. *Trends Plant Sci, 24*(11), 1022-1031. doi:10.1016/j.tplants.2019.07.007

Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., & Ben-Tal, N. (2016). ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res, 44*(W1), W344-350. doi:10.1093/nar/gkw408

Blount, Z. D., Lenski, R. E., & Losos, J. B. (2018). Contingency and determinism in evolution: Replaying life's tape. *Science, 362*(6415), eaam5979. doi:10.1126/science.aam5979

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

Buren, S., Jimenez-Vicente, E., Echavarri-Erasun, C., & Rubio, L. M. (2020). Biosynthesis of Nitrogenase Cofactors. *Chem Rev, 120*(12), 4921-4968. doi:10.1021/acs.chemrev.9b00489

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

Carruthers, B. M., Garcia, A. K., Rivier, A., & Kacar, B. (2021). Automated Laboratory Growth Assessment and Maintenance of *Azotobacter vinelandii*. *Curr Protoc Microbiol, 1*(3), e57. doi:10.1002/cpz1.57

Castle, S. D., Grierson, C. S., & Gorochowski, T. E. (2021). Towards an engineering theory of evolution. *Nat Commun, 12*(1), 3326. doi:10.1038/s41467-021-23573-3

Catling, D. C., & Zahnle, K. J. (2020). The Archean atmosphere. *Science Advances, 6*(9), eaax1420. doi:10.1126/sciadv.aax1420

Christiansen, J., Goodwin, P. J., Lanzilotta, W. N., Seefeldt, L. C., & Dean, D. R. (1998). Catalytic and biophysical properties of a nitrogenase Apo-MoFe protein produced by a *nifB*-deletion mutant of *Azotobacter vinelandii*. *Biochemistry, 37*(36), 12611-12623. doi:10.1021/bi981165b

Clifton, B. E., Kozome, D., & Laurino, P. (2022). Efficient Exploration of Sequence Space by Sequence-Guided Protein Engineering and Design. *Biochemistry*. doi:10.1021/acs.biochem.1c00757

Copley, S. D. (2021). Setting the stage for evolution of a new enzyme. *Curr Opin Struct Biol, 69*, 41-49. doi:10.1016/j.sbi.2021.03.001

Corbin, J. L. (1984). Liquid Chromatographic-Fluorescence Determination of Ammonia from Nitrogenase Reactions: A 2-Min Assay. *Applied and Environmental Microbiology, 47*(5), 1027-1030. doi:10.1128/aem.47.5.1027-1030.1984

Detroy, R. W., Witz, D. F., Parejko, R. A., & Wilson, P. W. (1968). Reduction of N_2 by complementary functioning of two components from nitrogen-fixing bacteria. *Proceedings of the National Academy of Sciences, 61*(2), 537-541. doi:10.1073/pnas.61.2.537

Dixon, R., & Kahn, D. (2004). Genetic regulation of biological nitrogen fixation. *Nat Rev Microbiol, 2*(8), 621-631. doi:10.1038/nrmicro954

Dos Santos, P. C. (2019). Genomic Manipulations of the Diazotroph *Azotobacter vinelandii*. *Methods Mol Biol, 1876*, 91-109. doi:10.1007/978-1-4939-8864-8_6

Dos Santos, P. C., Dean, D. R., Hu, Y., & Ribbe, M. W. (2004). Formation and insertion of the nitrogenase iron-molybdenum cofactor. *Chem Rev, 104*(2), 1159-1173. doi:10.1021/cr020608l

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

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

Eck, R. V., & Dayhoff, M. O. (1966). Evolution of the Structure of Ferredoxin Based on Living Relics of Primitive Amino Acid Sequences. *Science, 152*(3720), 363-366. doi:10.1126/science.152.3720.363

Falkowski, P. G. (1997). Evolution of the nitrogen cycle and its influence on the biological sequestration of CO_2 in the ocean. *Nature, 387*(6630), 272-275. doi:10.1038/387272a0

Ferrada, E., & Wagner, A. (2008). Protein robustness promotes evolutionary innovations on large evolutionary time-scales. *Proc Biol Sci, 275*(1643), 1595-1602. doi:10.1098/rspb.2007.1617

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

Gallon, J. R. (1992). Reconciling the Incompatible - N_2 Fixation and O_2. *New Phytologist, 122*(4), 571-609.

Garcia, A. K., Cavanaugh, C. M., & Kacar, B. (2021). The curious consistency of carbon biosignatures over billions of years of Earth-life coevolution. *ISME J*. doi:10.1038/s41396-021-00971-5

Garcia, A. K., & Kacar, B. (2019). How to resurrect ancestral proteins as proxies for ancient biogeochemistry. *Free Radic Biol Med, 140*, 260-269. doi:10.1016/j.freeradbiomed.2019.03.033

Garcia, A. K., McShea, H., Kolaczkowski, B., & Kacar, B. (2020). Reconstructing the evolutionary history of nitrogenases: Evidence for ancestral molybdenum-cofactor utilization. *Geobiology, 18*, 394-411. doi:10.1111/gbi.12381

Ghebreamlak, S. M., & Mansoorabadi, S. O. (2020). Divergent Members of the Nitrogenase Superfamily: Tetrapyrrole Biosynthesis and Beyond. *Chembiochem, 21*(12), 1723-1728. doi:10.1002/cbic.201900782

Gold, D. A., Caron, A., Fournier, G. P., & Summons, R. E. (2017). Paleoproterozoic sterol biosynthesis and the rise of oxygen. *Nature, 543*(7645), 420-423. doi:10.1038/nature21412

Goldford, J. E., Hartman, H., Smith, T. F., & Segre, D. (2017). Remnants of an ancient metabolism without phosphate. *Cell, 168*(6), 1126-1134. doi:10.1016/j.cell.2017.02.001

Hardy, R. W., Holsten, R. D., Jackson, E. K., & Burns, R. C. (1968). The acetylene-ethylene assay for N_2 fixation: laboratory and field evaluation. *Plant Physiol, 43*(8), 1185-1207. doi:10.1104/pp.43.8.1185

Harms, M. J., Eick, G. N., Goswami, D., Colucci, J. K., Griffin, P. R., Ortlund, E. A., & Thornton, J. W. (2013). Biophysical mechanisms for large-effect mutations in the evolution of steroid hormone receptors. *Proc Natl Acad Sci U S A, 110*(28), 11475-11480. doi:10.1073/pnas.1303930110

Harris, D. F., Lukoyanov, D. A., Kallas, H., Trncik, C., Yang, Z. Y., 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, 58*(30), 3293-3301. doi:10.1021/acs.biochem.9b00468

Harris, D. F., Lukoyanov, D. A., Shaw, S., Compton, P., Tokmina-Lukaszewska, M., Bothner, B., . . . Seefeldt, L. C. (2018). Mechanism of N_2 Reduction Catalyzed by Fe-Nitrogenase Involves Reductive Elimination of H_2. *Biochemistry, 57*(5), 701-710. doi:10.1021/acs.biochem.7b01142

Hochberg, G. K. A., & Thornton, J. W. (2017). Reconstructing ancient proteins to understand the causes of structure and function. *Annu Rev Biophys, 46*, 247-269. doi:10.1146/annurev-biophys-070816-033631

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

Hoffman, P. F., Kaufman, A. J., Halverson, G. P., & Schrag, D. P. (1998). A neoproterozoic snowball earth. *Science, 281*(5381), 1342-1346.

Hurley, S. J., Wing, B. A., Jasper, C. E., Hill, N. C., & Cameron, J. C. (2021). Carbon isotope evidence for the global physiology of Proterozoic cyanobacteria. *Science Advances, 7*(2). doi:10.1126/sciadv.abc8998

Inomura, K., Bragg, J., & Follows, M. J. (2017). A quantitative analysis of the direct and indirect costs of nitrogen fixation: a model based on *Azotobacter vinelandii*. *ISME J, 11*(1), 166-175. doi:10.1038/ismej.2016.97

Jablonski, D. (2004). Extinction: past and present. *Nature, 427*(6975), 589. doi:10.1038/427589a

Jemth, P., Karlsson, E., Vögeli, B., Guzovsky, B., Andersson, E., Hultqvist, G., . . . Chi, C. N. (2018). Structure and dynamics conspire in the evolution of affinity between intrinsically disordered proteins. *Science Advances, 4*(10). doi:10.1126/sciadv.aau4130

Jimenez-Vicente, E., Martin Del Campo, J. S., Yang, Z. Y., Cash, V. L., Dean, D. R., & Seefeldt, L. C. (2018). Application of affinity purification methods for analysis of the nitrogenase system from *Azotobacter vinelandii*. *Methods Enzymol, 613*, 231-255. doi:10.1016/bs.mie.2018.10.007

Kacar, B., Garmendia, E., Tuncbag, N., Andersson, D. I., & Hughes, D. (2017). Functional Constraints on Replacing an Essential Gene with Its Ancient and Modern Homologs. *MBio, 8*(4), e01276-01217. doi:10.1128/mBio.01276-17

Kacar, B., Ge, X., Sanyal, S., & Gaucher, E. A. (2017). Experimental Evolution of *Escherichia coli* Harboring an Ancient Translation Protein. *J Mol Evol, 84*(2-3), 69-84. doi:10.1007/s00239-017-9781-0

Kacar, B., Guy, L., Smith, E., & Baross, J. (2017). Resurrecting ancestral genes in bacteria to interpret ancient biosignatures. *Philos Trans A Math Phys Eng Sci, 375*(2109), 20160352. doi:10.1098/rsta.2016.0352

Kedzior, M., Garcia, A. K., Li, M., Taton, A., Adam, Z. R., Young, J. N., & Kacar, B. (2022). Resurrected Rubisco suggests uniform carbon isotope signatures over geologic time. *Cell Rep, 39*(4), 110726. doi:10.1016/j.celrep.2022.110726

Khadka, N., Dean, D. R., Smith, D., Hoffman, B. M., Raugei, S., & Seefeldt, L. C. (2016). CO_2 Reduction Catalyzed by Nitrogenase: Pathways to Formate, Carbon Monoxide, and Methane. *Inorg Chem, 55*(17), 8321-8330. doi:10.1021/acs.inorgchem.6b00388

Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., . . . Rives, A. (2022). Language models of protein sequences at the scale of evolution enable accurate structure prediction. *bioRxiv*. doi:10.1101/2022.07.20.500902

Love, G. D., Grosjean, E., Stalvies, C., Fike, D. A., Grotzinger, J. P., Bradley, A. S., . . . Summons, R. E. (2009). Fossil steroids record the appearance of Demospongiae during the Cryogenian period. *Nature, 457*(7230), 718-721. doi:10.1038/nature07673

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

McInnes, L., Healy, J., & Melville, J. (2020). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv*. doi:10.48550/arXiv.1802.03426

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

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

Navarro-Gonzalez, R., McKay, C. P., & Mvondo, D. N. (2001). A possible nitrogen crisis for Archaean life due to reduced nitrogen fixation by lightning. *Nature, 412*(6842), 61-64. doi:Doi 10.1038/35083537

Noar, J. D., & Bruno-Barcena, J. M. (2018). *Azotobacter vinelandii*: the source of 100 years of discoveries and many more to come. *Microbiology, 164*(4), 421-436. doi:10.1099/mic.0.000643

Ohno, S. (1970). *Evolution by gene duplication*. Berlin: Springer-Verlag.

Page, W. J., & von Tigerstrom, M. (1979). Optimal conditions for transformation of *Azotobacter vinelandii*. *Journal of Bacteriology, 139*(3), 1058-1061. doi:10.1128/jb.139.3.1058-1061.1979

Parsons, C., Stüeken, E. E., Rosen, C. J., Mateos, K., & Anderson, R. E. (2020). Radiation of nitrogen-metabolizing enzymes across the tree of life tracks environmental transitions in Earth history. *Geobiology, 19*, 18-34. doi:10.1111/gbi.12419

Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., . . . Ferrin, T. E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci, 30*(1), 70-82. doi:10.1002/pro.3943

Poza-Carrion, C., Jimenez-Vicente, E., Navarro-Rodriguez, M., Echavarri-Erasun, C., & Rubio, L. M. (2014). Kinetics of Nif gene expression in a nitrogen-fixing bacterium. *J Bacteriol, 196*(3), 595-603. doi:10.1128/JB.00942-13

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

Risso, V. A., Gavira, J. A., & Sanchez-Ruiz, J. M. (2014). Thermostable and promiscuous Precambrian proteins. *Environmental Microbiology, 16*(6), 1485-1489. doi:10.1111/1462-2920.12319

Rives, A., Meier, J., Sercu, T., Goyal, S., Lin, Z., Liu, J., . . . Fergus, R. (2021). Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proceedings of the National Academy of Sciences, 118*(15). doi:10.1073/pnas.2016239118

Sanchez-Baracaldo, P., & Cardona, T. (2020). On the origin of oxygenic photosynthesis and Cyanobacteria. *New Phytologist, 225*(4), 1440-1446. doi:10.1111/nph.16249

Sanchez-Baracaldo, P., Ridgwell, A., & Raven, J. A. (2014). A neoproterozoic transition in the marine nitrogen cycle. *Curr Biol, 24*(6), 652-657. doi:10.1016/j.cub.2014.01.041

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat Methods, 9*(7), 671-675. doi:10.1038/nmeth.2089

Schulz, L., Sendker, F. L., & Hochberg, G. K. A. (2022). Non-adaptive complexity and biochemical function. *Current Opinion in Structural Biology, 73*. doi:10.1016/j.sbi.2022.102339

Seefeldt, L. C., Yang, Z. Y., Lukoyanov, D. A., Harris, D. F., Dean, D. R., Raugei, S., & Hoffman, B. M. (2020). Reduction of Substrates by Nitrogenases. *Chem Rev, 120*, 5082-5106. doi:10.1021/acs.chemrev.9b00556

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

Shomar, H., & Bokinsky, G. (2021). Towards a Synthetic Biology Toolset for Metallocluster Enzymes in Biosynthetic Pathways: What We Know and What We Need. *Molecules, 26*(22). doi:10.3390/molecules26226930

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

Smith, B. E., Thorneley, R. N., Eady, R. R., & Mortenson, L. E. (1976). Nitrogenases from *Klebsiella pneumoniae* and *Clostridium pasteurianum*. Kinetic investigations of cross-

reactions as a probe of the enzyme mechanism. *Biochemical Journal, 157*(2), 439-447. doi:10.1042/bj1570439

Smith, M. J., & Francis, M. B. (2016). A Designed *A. vinelandii-S. elongatus* Coculture for Chemical Photoproduction from Air, Water, Phosphate, and Trace Metals. *ACS Synth Biol, 5*(9), 955-961. doi:10.1021/acssynbio.6b00107

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

Som, S. M., Buick, R., Hagadorn, J. W., Blake, T. S., Perreault, J. M., Harnmeijer, Jelte P., & Catling, D. C. (2016). Earth's air pressure 2.7 billion years ago constrained to less than half of modern levels. *Nature Geoscience, 9*(6), 448-451. doi:10.1038/ngeo2713

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

Spence, M. A., Kaczmarski, J. A., Saunders, J. W., & Jackson, C. J. (2021). Ancestral sequence reconstruction for protein engineers. *Curr Opin Struct Biol, 69*, 131-141. doi:10.1016/j.sbi.2021.04.001

Sprouffske, K., & Wagner, A. (2016). Growthcurver: an R package for obtaining interpretable metrics from microbial growth curves. *BMC Bioinformatics, 17*, 172. doi:10.1186/s12859-016-1016-7

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

Stripp, S. T., Duffus, B. R., Fourmond, V., Leger, C., Leimkuhler, S., Hirota, S., . . . Ribbe, M. W. (2022). Second and Outer Coordination Sphere Effects in Nitrogenase, Hydrogenase, Formate Dehydrogenase, and CO Dehydrogenase. *Chem Rev, 122*(14), 11900-11973. doi:10.1021/acs.chemrev.1c00914

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

Stüeken, E. E., Kipp, M. A., Koehler, M. C., & Buick, R. (2016). The evolution of Earth's biogeochemical nitrogen cycle. *Earth-Science Reviews, 160*, 220-239. doi:10.1016/j.earscirev.2016.07.007

Thompson, K. J., Kenward, P. A., Bauer, K. W., Warchola, T., Gauger, T., Martinez, R., . . . Crowe, S. A. (2019). Photoferrotrophy, deposition of banded iron formations, and methane production in Archean oceans. *Science Advances, 5*(11). doi:10.1126/sciadv.aav2869

Venkataram, S., Monasky, R., Sikaroodi, S. H., Kryazhimskiy, S., & Kacar, B. (2020). Evolutionary stalling and a limit on the power of natural selection to improve a cellular module. *Proc Natl Acad Sci U S A, 117*(31), 18582-18590. doi:10.1073/pnas.1921881117

Vicente, E. J., & Dean, D. R. (2017). Keeping the nitrogen-fixation dream alive. *Proc Natl Acad Sci U S A, 114*(12), 3009-3011. doi:10.1073/pnas.1701560114

Webb, B., & Sali, A. (2016). Comparative Protein Structure Modeling Using MODELLER. *Current Protocols in Bioinformatics, 54*(1). doi:10.1002/cpbi.3

Wolfe, J. M., & Fournier, G. P. (2018). Horizontal gene transfer constrains the timing of methanogen evolution. *Nat Ecol Evol, 2*(5), 897-903. doi:10.1038/s41559-018-0513-7

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., Dean, D. R., & Seefeldt, L. C. (2011). Molybdenum nitrogenase catalyzes the reduction and coupling of CO to form hydrocarbons. *J Biol Chem, 286*(22), 19417-19421. doi:10.1074/jbc.M111.229344

Zehr, J. P., Jenkins, B. D., Short, S. M., & Steward, G. F. (2003). Nitrogenase gene diversity and microbial community structure: a cross-system comparison. *Environmental Microbiology, 5*(7), 539-554. doi:10.1046/j.1462-2920.2003.00451.x

Zerkle, A. L., House, C. H., Cox, R. P., & Canfield, D. E. (2006). Metal limitation of cyanobacterial N_2 fixation and implications for the Precambrian nitrogen cycle. *Geobiology, 4*(4), 285-297. doi:10.1111/j.1472-4669.2006.00082.x