# Conservation of nitrogenase functionality over long timescales

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

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

[b] NASA Center for Early Life and Evolution, University of Wisconsin-Madison, WI 53706

[c] Lunar and Planetary Laboratory, University of Arizona, Tucson, AZ, 85711

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

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

**Classification:** Major–Biological Sciences; Dual minor–Evolution/Systems Biology

**Keywords: protein evolution, metabolic engineering, nitrogenase, nitrogen fixation, paleogenetics**

## ABSTRACT

Understanding the coevolutionary dynamics between proteins, cellular networks, and environmental pressures remains an outstanding challenge in biology. The effects of molecular constraints on the development of metabolic pathways that shape the biosphere remain underexplored. Extant genes, relics of a >3-billion-year history of life, offer the means to experimentally reconstruct protein evolutionary trajectories. Surveying life's extinct genetic repertoire may unlock adaptive states to environmental conditions unobserved today. However, the lack of suitable experimental systems poses a hurdle for such strategies, particularly for the reconstruction of biogeochemically critical ancient proteins and their supporting networks. To fill this gap, here we developed an evolutionarily guided, systems biology and biochemistry approach for the resurrection and functional characterization of ancient nitrogen fixation machinery in *Azotobacter vinelandii*. We report the recovery of active, ancestral nitrogenase enzymes and observe the conservation of core, catalytic features that are robust to numerous residue-level changes and modular incorporation of ancestral protein components. These results highlight the preservation of protein features vital for primary function over long timescales. An ancient-modern hybrid experimental strategy enables the reconstruction and historical characterization of essential biosystems, providing an evolutionarily vetted toolbox for the study, resurrection, and engineering of life's key metabolic innovations.

## INTRODUCTION

Metabolic proteins are primary interfaces between processes occurring within cells and in the environment. The evolutionary histories of these proteins can be imprinted by environmental and biological drivers of metabolic innovations. Deconvoluting these drivers is particularly challenging when such proteins are formed of multicomponent complexes, participate in a broader protein interaction network, are dependent on sophisticated cofactors for activity, or are inextricably linked with planetary-scale transitions (1-5). The latter factor grants a venue to study how novel protein functionality arises and persists (6-8), and how emergent novelty in protein structure and interaction networks (8-11) can come to govern the planetary biosphere. To explore these factors a systems biology approach that considers ancient protein functionality in the context of its interacting partners, genetic regulation, and cellular context is necessary (12).

The study of biological nitrogen fixation presents such an opportunity. By all available evidence, the origin of nitrogen fixation was likely a singular evolutionary event on which the modern biosphere has since been built. Nitrogen fixation is catalyzed by an early evolved (13, 14) and conserved family of metalloenzymes called nitrogenases that represent the sole remaining enzymatic solution for obtaining bioavailable ammonia ($NH_3$) from highly inert, atmospheric dinitrogen ($N_2$). Though fundamentally reliant on this single enzyme for the $N_2$ reduction step, nitrogen fixation in modern prokaryotes typically requires numerous genes that encode the nitrogenase structural components themselves, in addition to the proteins and regulatory synchronicity required for their assembly and stability (15-17). How nitrogenases and their supporting networks emerged and co-evolved under past environmental parameters is still poorly constrained relative to their importance in Earth's planetary history. Because nitrogen has been a limiting nutrient over geological timescales (18-20), nitrogenase functionality 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 costly workaround for nitrogen fertilizer production (21) designed to supplement a remarkable molecular innovation that biology has tinkered with for more than 3 billion years (13, 14).

To study the evolution of nitrogenase function over time, we established an evolutionary metabolic engineering approach for the characterization of phylogenetically inferred ancestors, both at the systems- and molecular-level. We report the laboratory resurrection of ancestral nitrogenase enzymes by the genomic incorporation of synthetic, ancient genes into the model nitrogen-fixing bacterium, *Azotobacter vinelandii* (*A. vinelandii*). We developed a "hybrid" strategy that modularly integrates synthetic ancestral genes into an extant microbe with a compatible genetic context. Variably replacing different protein subunits of the nitrogenase complex with ancestral counterparts complements nitrogen fixation in *A. vinelandii*. Further biochemical characterization of purified ancient nitrogenases in this evolutionary framework show that ancestral enzymes retain the core $N_2$ reduction mechanism of their extant counterparts and maintain the same catalytic selectivity between $N_2$ and protons. Our results highlight the robustness of primary enzyme functionality to historical amino acid substitutions reconstructed over long timescales.

## RESULTS

### A metabolic resurrection strategy for investigation of ancestral nitrogenase function

The genomic engineering and functional assessment of ancestral nitrogenases required a suitable diazotrophic microbial host, owing to the challenge associated with heterologous expression of these proteins (12, 21, 22). We selected the obligately aerobic gammaproteobacterium, *A. vinelandii*, 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 (23-25).

Extant *A. vinelandii* molybdenum-dependent ("Mo-") nitrogenase is among the primary targets of study for insights into nitrogenase biochemical, structural, and regulatory insights (26-28). The wild-type (WT) Mo-nitrogenase complex comprises multiple unique subunits and thus multiple potential protein targets for ancestral reconstruction. These subunits, NifH, NifD, and NifK, are arranged into two catalytic components: a $NifH_2$ homodimer and a $NifD_2K_2$ heterotetramer. During each catalytic cycle, both components transiently associate to transfer one electron from $NifH_2$ to $NifD_2K_2$ and subsequently dissociate. Electrons are delivered and accumulate at the active-site metallocluster ("FeMoco") housed within the NifD subunits for reduction of $N_2$ substrate to $NH_3$. From a nitrogenase protein phylogeny (see Materials and Methods), we identified an evolutionary lineage ancestral to extant *A. vinelandii* WT nitrogenase, providing several candidate ancestral proteins of varying sequence similarity to their modern counterparts (**Fig. 1A**). Two nodes within this lineage, "Anc1" and "Anc2" (in order of increasing age), were targeted to reconstruct nitrogenase ancestors with as little as 83% protein sequence identity to WT (**Table S1**).

For reconstruction of Anc1 proteins, we designed and executed two parallel experimental strategies (**Fig. 1B**). First, we evaluated the tolerance of modern *A. vinelandii* to ancestral nitrogenase subunits of increasing age, generating enzyme hybrids that harbor ancestral, FeMoco-containing NifD subunits ("$NifD^{Anc1}$ and $NifD^{Anc2}$") within an otherwise modern nitrogenase complex. Second, we explored whether this $N_2$-fixing microbial model could accommodate multiple ancestral nitrogenase subunits by reconstructing all three ancestral NifHDK proteins ("$NifHDK^{Anc1}$") in a single complex (**Fig. 1B**). For both Anc1 and Anc2, we considered only the most probable ancestral subunit sequences, or the sequences that maximize the posterior probability of each possible ancestral amino acid at each protein site (see Materials and Methods). Mean posterior probabilities for all targeted subunit sequences range between 0.95 and 0.99. Approximately 1-3% of the amino acid sites of any of the investigated subunit sequences had plausible alternate ancestral states and, of these, only three sites are phylogenetically conserved (see Materials and Methods; **Table S1**).

Residue-level differences between ancestral and WT nitrogenases are broadly distributed along the length of each ancestral sequence (**Fig. 2A**). Protein sequence identities of each ancestral subunit relative to WT range from ~83% to 94% (**Table S1**). Residue-level differences are predominantly relegated to the "outer shell" of the nitrogenase structure where there is relatively low conservation among taxonomically diverse nitrogenases. The sole ancestral substitution proximal to FeMoco is NifD I355V (observed in all targeted NifD ancestors; residue numbering here and hereafter from *A. vinelandii* Mo-nitrogenase), a site belonging to a loop considered important for FeMoco insertion (29) (**Fig. 2B**). However, this variation is not unexpected given the presence of valine at this site in the majority (~88%) of Mo-nitrogenases in our phylogenetic dataset. Other ancestral substitutions are notable for their location at phylogenetically conserved residue sites and/or within subunit interfaces, including two proximal substitutions at the NifD:NifK interface, F429Y (NifD) and R108K (NifK) (**Fig. 2C**). In all studied ancestors, the C275 and H442 FeMoco ligands as well as other strictly conserved nitrogenase residues are retained.

Nucleotide sequences for ancestral nitrogenase proteins were codon-optimized for expression in *A. vinelandii*, synthesized, and cloned into plasmids for subsequent genomic engineering and protein purification (see Materials and Methods). We leveraged a two-step strategy for markerless integration of ancestral nitrogenase genes into the *A. vinelandii* genome, following Dos Santos (24). Native *nifD* or *nifHDK* genes were first knocked out by a selective marker followed by replacement of the marker with the synthetic ancient gene(s), yielding three engineered strains: AK013, AK023, and AK014. Descriptions of the strains are listed in **Table 1**.

**Ancestral nitrogenases complement *in vivo* nitrogen fixation**

All *A. vinelandii* strains engineered with ancestral *nif* genes were capable of diazotrophic growth in molybdenum-containing, nitrogen-free media. Mean growth rates for all strains were comparable to WT ($p > .05$), though we observed a ~24-h increase in the lag phase of strain AK014 (**Fig. 3A, Supplementary Table S2**). Cultures of the Δ*nifD* strain DJ2278 (**Table 1**; Materials and Methods) prepared in similar conditions did not yield detectable growth, indicating that the growth observed for ancestral strains was unlikely to stem from leaky expression of the alternative, vanadium- ("V-") or iron- ("Fe-") dependent nitrogen fixation genes in *A. vinelandii* (31). *In vivo* ancestral nitrogenase activity was quantified by the reduction rate of the non-physiological substrate, acetylene ($C_2H_2$) (32). We found that WT and AK023 exhibited

comparable mean acetylene reduction rates, whereas rates for both AK013 and AK014 were ~50% that of WT ($p < .01$) (**Fig. 3B**).

*A. vinelandii* strains were grown at scale for purification of ancestral (AK023) or ancestral-extant hybrid (AK013, AK014) $NifD_2K_2$ proteins and subsequent biochemical analysis (see Materials and Methods). Previously studied Mo-nitrogenase variants as well as alternative V- and Fe-nitrogenases exhibit varying reactivities to a number of different substrates (26), reflecting the contribution of both metallocluster composition and polypeptide environment to catalytic properties. Here, purified $NifD_2K_2$ variants coupled to WT $NifH_2$ were assayed for reduction of $N_2$ as well as protons and $C_2H_2$. Ancestral nitrogenases reduced all three substrates at slower rates relative to WT, in part possibly attributable to poor cofactor occupancy or lability of the isolated protein (**Fig. 3C**).

**Ancestral nitrogenases exhibit a conserved binding mechanism and efficiency for $N_2$ reduction**

Recent work has demonstrated that extant nitrogenase isozymes (Mo-, V-, Fe-) follow the same general mechanism for $N_2$ binding and reduction (26, 33). The mechanism involves the accumulation of 4 electron/protons on the active-site cofactor as metal bound hydrides to generate the $E_4(4H)$ state (**Fig. 2D**). 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 ($H_2$) (**Fig. 2D**). $N_2$ binding is reversible in the presence of sufficient $H_2$, which displaces $N_2$ 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 (measured for purified $NifD_2K_2$ variants coupled to WT $NifH_2$) (**Fig. 3D**).

In the event the $E_4(4H)$ state fails to capture $N_2$, nitrogenases will simply evolve $H_2$ from the $E_4(4H)$ state to generate the $E_2(2H)$ state (**Fig. 2D**). The ratio of $H_2$ evolved to $N_2$ reduced ($H_2/N_2$) can be used as a measure of the performance, or efficiency, of nitrogenases in utilizing 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. $H_2/N_2$ values for all ancestors under 1 atm $N_2$ was ~2, similar to

extant Mo-nitrogenase (**Fig. 2D**). Together, both $N_2$ binding mechanism and efficiencies across resurrected ancestral nitrogenase proteins are notably conserved, despite numerous residue-level changes.

## DISCUSSION

Resurrected ancestral nitrogenases can effectively reduce dinitrogen and acetylene, both within modern *A. vinelandii* cells and in purified form. All engineered strains recover growth rates in diazotrophic conditions comparable to WT. $H_2$ inhibits the reduction of $N_2$ in purified ancestors and $H_2/N_2$ ratios are within the range characteristic of extant Mo-nitrogenase. The mechanism for $N_2$ binding is shared by all extant isozymes (33) as well as the catalytic efficiency observed for modern Mo-nitrogenase is thus preserved in the studied ancestors. These properties are maintained despite substantial residue-level changes to peripheral regions of the nitrogenase complex in addition to select ancestral substitutions within more conserved, active-site or protein-interface regions.

Kinetic analyses 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 in an engineered ancestral-hybrid metabolism. A striking feature of our results is the comparison between the phenotypic properties of nitrogenase complexes containing single versus multiple ancestral protein subunits (i.e., $NifD^{Anc1}$ in AK013 versus $NifHDK^{Anc1}$ in AK023). The incorporation of ancestral NifH and NifK subunits appears to have neutral or even beneficial phenotypic effects beyond that observed for enzymes that only incorporate ancestral NifD. Growth rates of relevant engineered strains are comparable, and we observe improved *in vivo* nitrogenase activities for strains containing all ancestral nitrogenase proteins. The results of our modular incorporation of ancestral, nitrogenase subunit proteins supports the greater robustness of nitrogenase functionality to changes in the NifH and NifK proteins compared to those in NifD. Greater sequence-level constraints for NifD may be expected due to its central role in housing the site of $N_2$ reduction.

Previous efforts to manipulate the molecular characteristics of nitrogenase enzymes have typically targeted the catalytic impacts of single or dual substitutions, often resulting in reduction or abolishment of nitrogenase activity (34, 35) though in certain cases improving reactivity toward

alternate, industrially relevant substrates (26, 36). Despite illuminating key features of modern nitrogenase catalysis, the combination of detailed functional studies and evolutionary context afforded by paleogenetic approaches have not previously been accomplished for the nitrogen fixation system. Future work may expand on the identification of nitrogen fixation components that preserve primary function—potentially targeting the circumstances under which relevant nitrogenase subunit interfaces emerged and were maintained (10)—as well as those that might be most amenable to manipulation for desired outcomes.

Attempts to explore protein sequence-function space via rational design can be hindered by local adaptive valleys. However, evolutionarily guided sequence exploration provides an effective strategy to efficiently traverse functional sequence space already "vetted" by natural selection and can integrate the effects of epistatic interactions that are challenging to predict a priori (8, 37, 38). A historical, synthetic biology approach focused within the lineage of the utilized model host system is perhaps more likely to yield a functional phenotype because introduced changes are constrained to the evolutionary "vernacular" of native proteins and their interaction networks (39). Such functional variation should theoretically have arisen from an overlapping suite of selective pressures, in contrast to those of extant orthologs descended from divergent evolutionary paths. The resurrection strategy leveraged here importantly provides a means to directly test paradigms of early protein and metabolic evolution, including those associated with the emergence, robustness, and malleability of protein functionality (6, 7, 40-42), the development of protein-protein interactions and complexes (9-11), and the role of evolutionary contingency (43-45). Nevertheless, the experimental application of ancestral sequence resurrection toward the systems-level investigation of earliest evolved, biogeochemically crucial metabolic pathways remains in its infancy, despite the potential for provocative connections between molecular-scale innovations, planetary-scale evolution, and independent, geological records of life (46, 47).

Our results demonstrate the tractability of leveraging phylogenetic models to carry out extensive, empirical manipulations of challenging enzymatic systems and their microbial hosts. This approach enables a means to introduce tens or hundreds of amino acid substitutions simultaneously while preserving core functionality. Seeking guidance from the Earth's environmental past as a creative force for design as well as manipulating whole biosystems has the potential to catalyze an entirely new area of bioengineering and evolutionary research.

## MATERIALS AND METHODS

### Nitrogenase ancestral sequence reconstruction and analysis

A nitrogenase protein sequence dataset was assembled by BLAST search of the NCBI non-redundant protein database (accessed August 2020), followed by manual curation 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 reductase sequences (Bch/ChlLNB), resulting in a 770-sequence dataset. For sequences used to construct strains AK013 and AK014, tree and marginal ancestral sequence inference were performed consecutively by RAxML v8.2.10 (48) using the LG+G+F evolutionary model. For strain AK023, tree inference was repeated in RAxML and ancestral sequences were reconstructed by PAML v4.9j (49) also using the LG+G+F model. Thus, two alternative nitrogenase phylogenies were incorporated in this study. Ancestral nitrogenase sequences were analyzed by identifying amino acid substitutions at phylogenetically conserved sites, defined by a site-wise conservation score >7 analyzed by the ConSurf server (50) (given an extant Mo-nitrogenase alignment). Structural homology models of ancestral sequences were generated by MODELLER v10.2 (51) using PDB 1M34 as a template for all nitrogenase protein subunits, and visualized by ChimeraX v1.3 (52).

### *A. vinelandii* strain engineering and culturing

Nucleotide sequences of ancestral nitrogenase proteins were codon-optimized for *A. vinelandii* expression, followed by synthesis and cloning into pUC19 vectors (Twist Bioscience, CA, USA; GenScript, NJ, USA). Inserts were designed with 400 base-pairs of homology to direct recombination at the *A. vinelandii nif* locus. An “ASWSHPQFEK” Strep-tag was included in at the N-terminal of each synthetic *nifD* gene for $NifD_2K_2$ affinity purification. See **Supplementary Table S3** for a list of plasmids used in this study.

Engineering of *A. vinelandii* strains followed Dos Santos (24). *A. vinelandii* WT and DJ2278 (Δ*nifD::*KmR) strains were generously provided by Dennis Dean (Virginia Tech). For generation of strains AK013 and AK014, DJ2278 was used as a parent strain (**Table 1**). An additional Δ*nifHDK::*KmR parent strain, AK022, was constructed from WT in this study for the generation of AK023 harboring ancestral *nifHDK*. See **Supplementary Table S3** for a list of strains used in this

study. Genetic competency was induced by subculturing parent strains in Mo- and Fe-deficient media. Competent cells were transformed with 1 µg of donor plasmid. Transformants were screened for rescue of diazotrophic phenotype and loss of kanamycin resistance, followed by Sanger sequencing of the *nifHDK* cluster. Transformants were passaged at least 3 times to ensure phenotypic stability prior to storage and additional characterization.

*A. vinelandii* strains were cultured in flasks containing Burk's medium (with modifications as needed) at 30 °C and shaking at 300 rpm. For growth rate quantification, seed cultures representing independent biological replicates were grown in nitrogen-supplemented Burk's medium (13 mM ammonium acetate) followed by inoculation in 50 mL nitrogen-free Burk's medium to an optical density of 0.01 at 600 nm (OD600), after Carruthers et al. (53). Optical densities under diazotrophic conditions were measured over 72 h. Growth parameters were estimated using the R package growthcurver (54) and statistically analyzed by one-way ANOVA and a Tukey HSD post-hoc test.

**_In vivo_ 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 under diazotrophic conditions to an OD600 ~ 0.5, at which point a rubber septum cap was affixed to the mouth of each flask. 50% of the flask headspace (25 mL) was removed and replaced by injecting an equivalent volume of acetylene gas. The cultures were subsequently shaken at 30 °C, 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, Kyoto, Japan). 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 for 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, NC, USA). Acetylene reduction rates for each replicate were normalized to total protein and statistically analyzed by one-way ANOVA and a Tukey HSD post-hoc test.

**Nitrogenase purification and biochemical characterization**

Ancient nitrogenase proteins were expressed from relevant *A. vinelandii* strains (AK013, AK014, AK023) and purified according to previously published methods with some modifications (55) . Cells were grown under diazotrophic conditions (i.e., no exogenously added fixed nitrogen source) and no derepression step to a sufficient OD600 before harvesting. WT $NifH_2$ was expressed in *A. vinelandii* strain DJ884 and purified by previously published methods (56). Protein purity was assessed at ≥ 95% by SDS-PAGE gel with Coomassie blue staining.

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 ($N_2$, $C_2H_2$, $H_2$) added to appropriate concentrations with any remaining balance of the headspace being argon. $NifD_2K_2$ proteins (~240 kDa) were added to 0.42 µM, the vial vented to atmospheric pressure, and the reaction initiated by addition of $NifH_2$ (~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 with some modifications (57). 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 472 nm and $NH_3$ quantified using a standard generated with $NH_4Cl$. $H_2$ and $C_2H_4$ were quantitated by gas chromatography with thermal conductivity detector (GC-TCD) and gas chromatography with flame ionization detector (GC-FID) respectively, according to published methods (58, 59).

## 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 two *A. vinelandii* strains and 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. We thank Joanna Masel 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; A.K.G, L.C.S., B.K.), and the UW-Madison College of Agricultural and Life Sciences, the NASA Postdoctoral Program (A.K.G.), the University of Arizona Foundation Small Grants Program (A.K.G., B.K.), NASA Arizona Space Grant (B.M.C.), and the NASA Early Career Faculty Award (B.K.).

## REFERENCES

1. P. G. Falkowski, T. Fenchel, E. F. Delong, The microbial engines that drive Earth's biogeochemical cycles. *Science* **320**, 1034-1039 (2008). https://doi.org/10.1126/science.1153213
2. J. Barnabas, R. M. Schwartz, M. O. Dayhoff, Evolution of major metabolic innovations in the Precambrian. *Origins of Life* **12**, 81-91 (1982). https://doi.org/10.1007/bf00926914
3. R. E. Blankenship, H. Hartman, The origin and evolution of oxygenic photosynthesis. *Trends Biochem Sci* **23**, 94-97 (1998). https://doi.org/10.1016/S0968-0004(98)01186-4
4. A. H. Knoll, M. A. Nowak, The timetable of evolution. http://dx.doi.org/10.1126/sciadv.1603076.
5. J. E. Goldford, D. Segrè, Modern views of ancient metabolic networks. *Current Opinion in Systems Biology* **8**, 117-124 (2018). https://doi.org/10.1016/j.coisb.2018.01.004
6. S. Ohno, *Evolution by gene duplication* (Springer-Verlag, Berlin, 1970), pp. 160.
7. S. D. Copley, Setting the stage for evolution of a new enzyme. *Curr Opin Struct Biol* **69**, 41-49 (2021). https://doi.org/10.1016/j.sbi.2021.03.001
8. G. K. A. Hochberg, J. W. Thornton, Reconstructing ancient proteins to understand the causes of structure and function. *Annu Rev Biophys* **46**, 247-269 (2017). https://doi.org/10.1146/annurev-biophys-070816-033631
9. P. Jemth *et al.*, Structure and dynamics conspire in the evolution of affinity between intrinsically disordered proteins. *Sci Adv* **4** (2018). https://doi.org/10.1126/sciadv.aau4130
10. L. Schulz, F. L. Sendker, G. K. A. Hochberg, Non-adaptive complexity and biochemical function. *Current Opinion in Structural Biology* **73** (2022). https://doi.org/10.1016/j.sbi.2022.102339
11. R. V. Eck, M. O. Dayhoff, Evolution of the Structure of Ferredoxin Based on Living Relics of Primitive Amino Acid Sequences. *Science* **152**, 363-366 (1966). https://doi.org/10.1126/science.152.3720.363

12. B. Kacar, L. Guy, E. Smith, J. Baross, Resurrecting ancestral genes in bacteria to interpret ancient biosignatures. *Philos Trans R Soc A* **375** (2017). https://doi.org/10.1098/rsta.2016.0352

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

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

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

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

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

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

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

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

21. E. J. Vicente, D. R. Dean, Keeping the nitrogen-fixation dream alive. *Proc Natl Acad Sci USA* **114**, 3009-3011 (2017). https://doi.org/10.1073/pnas.1701560114

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

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

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

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

26. L. C. Seefeldt *et al.*, Reduction of Substrates by Nitrogenases. *Chem Rev* **120**, 5082-5106 (2020). https://doi.org/10.1021/acs.chemrev.9b00556

27. T. Spatzal *et al.*, Evidence for interstitial carbon in nitrogenase FeMo cofactor. *Science* **334**, 940 (2011). https://doi.org/10.1126/science.1214025

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

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

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

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

32. R. W. Hardy, R. D. Holsten, E. K. Jackson, R. C. Burns, The acetylene-ethylene assay for n(2) fixation: laboratory and field evaluation. *Plant Physiol* **43**, 1185-1207 (1968). https://doi.org/10.1104/pp.43.8.1185

33. D. F. Harris *et al.*, Mo-, V-, and Fe-Nitrogenases Use a Universal Eight-Electron Reductive-Elimination Mechanism To Achieve N2 Reduction. *Biochemistry* **58**, 3293-3301 (2019). https://doi.org/10.1021/acs.biochem.9b00468

34. B. M. Barney, R. Y. Igarashi, P. C. Dos Santos, D. R. Dean, L. C. Seefeldt, Substrate interaction at an iron-sulfur face of the FeMo-cofactor during nitrogenase catalysis. *J Biol Chem* **279**, 53621-53624 (2004). https://doi.org/10.1074/jbc.M410247200

35. K. E. Brigle *et al.*, Site-directed mutagenesis of the nitrogenase MoFe protein of Azotobacter vinelandii. *Proc Natl Acad Sci USA* **84**, 7066-7069 (1987).

36. K. R. Fixen *et al.*, Light-driven carbon dioxide reduction to methane by nitrogenase in a photosynthetic bacterium. *Proc Natl Acad Sci USA* **113**, 10163-10167 (2016). https://doi.org/10.1073/pnas.1611043113

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

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

39. B. Kacar, "Rolling the Dice Twice: Evolving Reconstructed Ancient Proteins in Extant Organisms" in *Chance in Evolution,* G. Ramsey, C. H. Pence, Eds. (University of Chicago Press, Chicago, IL, 2016), pp. 264-276.

40. M. J. Harms *et al.*, Biophysical mechanisms for large-effect mutations in the evolution of steroid hormone receptors. *Proc Natl Acad Sci USA* **110**, 11475-11480 (2013). https://doi.org/10.1073/pnas.1303930110

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

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

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

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

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

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

47. M. Kedzior *et al.*, Resurrected Rubisco suggests uniform carbon isotope signatures over geologic time. *Cell Rep* **39**, 110726 (2022). https://doi.org/10.1016/j.celrep.2022.110726

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

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

50. H. Ashkenazy *et al.*, ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res* **44**, W344-350 (2016). https://doi.org/10.1093/nar/gkw408

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

52. E. F. Pettersen *et al.*, UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci* **30**, 70-82 (2021). https://doi.org/10.1002/pro.3943

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

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

55. E. Jimenez-Vicente *et al.*, Application of affinity purification methods for analysis of the nitrogenase system from Azotobacter vinelandii. *Methods Enzymol* **613**, 231-255 (2018). https://doi.org/10.1016/bs.mie.2018.10.007

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

57. J. L. Corbin, Liquid Chromatographic-Fluorescence Determination of Ammonia from Nitrogenase Reactions: A 2-Min Assay. *Appl Environ Microbiol* **47**, 1027-1030 (1984). https://doi.org/10.1128/aem.47.5.1027-1030.1984

58. N. Khadka *et al.*, $CO_2$ Reduction Catalyzed by Nitrogenase: Pathways to Formate, Carbon Monoxide, and Methane. *Inorg Chem* **55**, 8321-8330 (2016). https://doi.org/10.1021/acs.inorgchem.6b00388

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

503 **TABLES**

504

505 **Table 1.** *A. vinelandii* strains used in this study. “Nif+”: diazotrophic, “Nif-“: non-diazotrophic.

| **Strain** | **Description** |
|---|---|
| DJ | Wild-type; Nif+ |
| DJ2278 | Δ*nifD*::KmR; Nif- |
| AK022 | Δ*nifHDK*::KmR; Nif- |
| AK013 | Δ*nifD*::*nifD*$^{Anc1}$; Nif+ |
| AK023 | Δ*nifHDK*::*nifHDK*$^{Anc1}$; Nif+ |
| AK014 | Δ*nifD*::*nifD*$^{Anc2}$; Nif+ |

506

## FIGURES AND FIGURE CAPTIONS

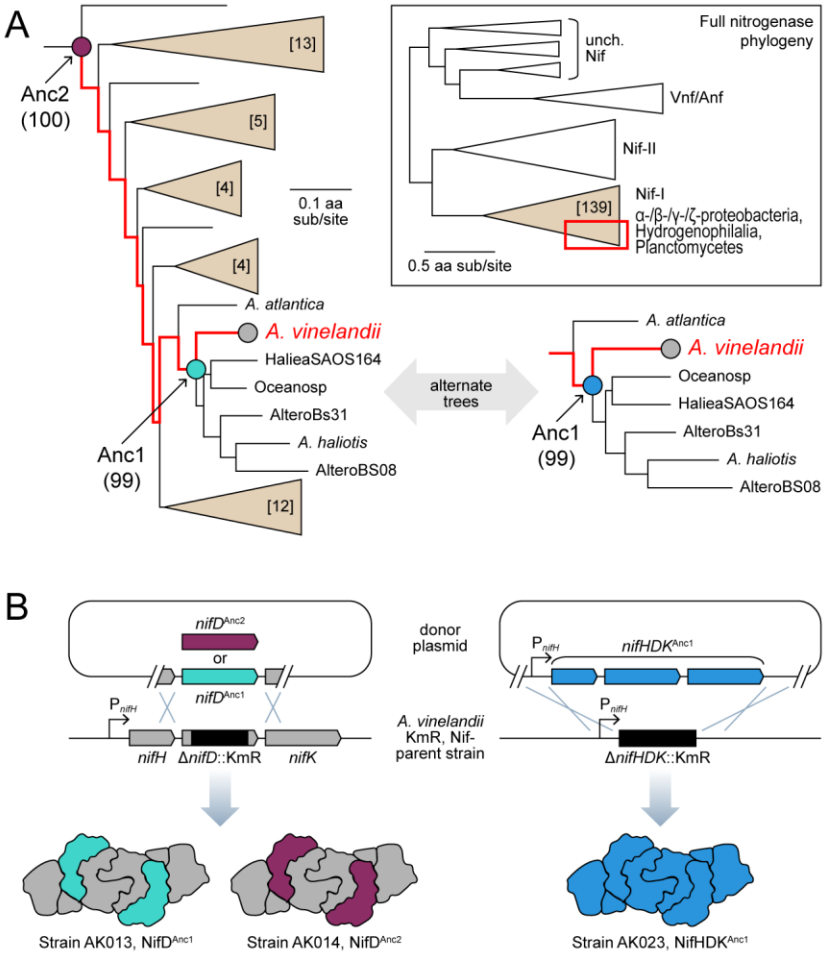


**Fig. 1.** Metabolic engineering strategy for the experimental resurrection of ancestral nitrogenases. (*A*) Phylogenetic subtrees constructed for the inference of ancestral nitrogenase protein sequences. Select nodes (Anc1, Anc2) within the *A. vinelandii* lineage (red) were targeted for gene synthesis. Analogous nodes for Anc1 are highlighted in each of two alternate tree reconstructions (see Materials and Methods). Full nitrogenase phylogeny is boxed and location of the subtrees is indicated by a red rectangle (nomenclature of the major nitrogenase clades follows Garcia et al. (30)). Number of sequences in collapsed clades are indicated by brackets and SH-like aLRT branch support values for selected nodes are displayed in parentheses. (*B*) *A. vinelandii g*enomic incorporation of synthetic ancestral nitrogenase genes. Kanamycin-resistant ("KmR"), non-diazotrophic ("Nif-") parent strains were used both for integration of ancestral *nifD* (left) and *nifHDK* (right) variants. Ancestral subunits in the illustrated, assembled nitrogenase complexes are colored teal, purple, or blue for each strategy (WT subunits shown in grey).

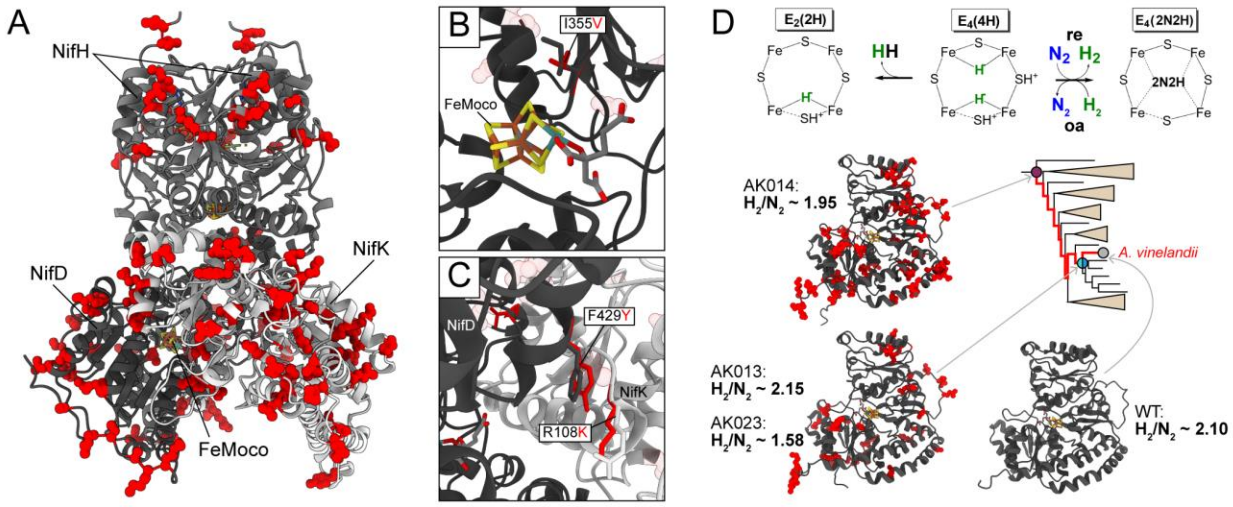


**Fig. 2.** Structural and sequence variation versus mechanistic conservation of ancestral nitrogenases. (*A*) Homology model of ancestral nitrogenase NifHDK protein structure (strain AK023, node Anc1; PDB 1M34 used as template). One catalytic half of the full nitrogenase complex is shown. Substitutions relative to WT are highlighted in red. (*B,C*) Select ancestral substitutions (strain AK023, node Anc1) relative to WT at relatively conserved protein sites (see Materials and Methods). Ancestral substitution side chains are highlighted in red and superimposed on those of WT (grey). Depictions use the same structural model shown in (*A*). (*B*) Ancestral peptide environment of FeMoco includes a I355V substitution. (*C*) Ancestral NifD:NifK interface. Two proximal mutations of ancestral NifD (F429Y) and NifK (R108K) proteins. (*D*) Experimental evidence that ancestral nitrogenases conserve the extant $N_2$-reduction mechanism. A partial schematic of the extant catalytic mechanism is shown above, centering on the $N_2$ binding $E_4$(4H) state of FeMo-co, which has accumulated 4 electrons/protons (see text for discussion). Ratios of evolved $H_2$ to reduced $N_2$ ($H_2/N_2$) for purified nitrogenase ancestors are similar to those of WT (~2), reflecting a conserved $N_2$ binding mechanism and efficiency. Ancestral amino acid substitutions in NifD subunits relative to WT are highlighted in red.

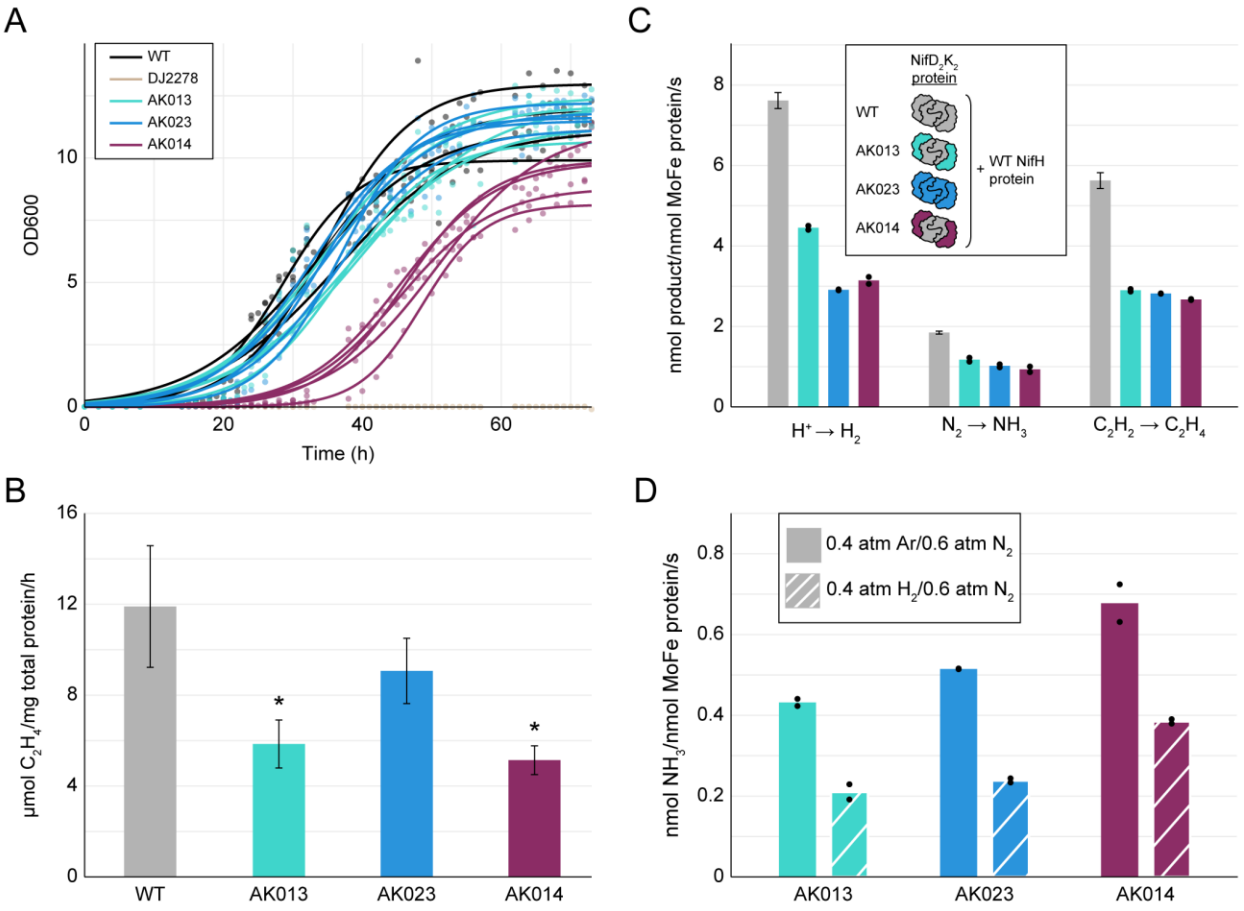


**Fig. 3.** Systems- and molecular-level characterization of WT and ancestral nitrogenases. (*A*) Growth curves of *A. vinelandii* strains measured by the optical density at 600 nm ("$OD_{600}$"). The fitted logistic curve for each of five biological replicates per strain is shown alongside raw data points (circles). The non-diazotrophic DJ2278 ( Δ*nifD*) strain was used as a negative control. (*B*) *In vivo* acetylene reduction rates quantified by production of ethylene ($C_2H_4$). Bars represent the mean of three biological replicates per strain and error bars indicate ± 1 SD. Asterisks indicate $p < .01$ compared to WT. (*C*) Specific activities of purified $NifD_2K_2$ proteins assayed with WT NifH protein. Bars represent the mean of either six (for WT) or two (all other strains) independent experiments. Error bars for WT indicate ± 1 SD, whereas individual data points for other strains are shown with black circles. (*D*) Inhibition of $N_2$ reduction by $H_2$. Bars represent the mean of two independent experiments and individual data points are shown in black circles. Assay was run using the same nitrogenase protein components as (*C*).