ARTICLE

Reconstruction of nitrogenase predecessors suggests origin from maturase-like enzymes

Amanda K. Garcia[1], Bryan Kolaczkowski[2], Betül Kaçar[1,3]*

[1]Department of Molecular and Cellular Biology, University of Arizona, Tucson, Arizona, USA;
[2]Department of Microbiology and Cell Science, University of Florida, Gainesville, Florida, USA;
[3]Department of Planetary Sciences, Lunar and Planetary Laboratory, University of Arizona, Tucson, Arizona, USA

*Corresponding author
Email: betul@arizona.edu

ABSTRACT

The evolution of biological nitrogen fixation, uniquely catalyzed by nitrogenase enzymes, has been one of the most consequential biogeochemical innovations over life's history. Though understanding the early evolution of nitrogen fixation has been a longstanding goal from molecular, biogeochemical, and planetary perspectives, its origins remain enigmatic. In this study, we reconstructed the evolutionary histories of nitrogenases, as well as homologous maturase proteins that participate in the assembly of the nitrogenase active-site cofactor but are not able to fix nitrogen. We combined phylogenetic and ancestral sequence inference with an analysis of predicted functionally divergent sites between nitrogenases and maturases to infer the nitrogen-fixing capabilities of their shared ancestors. Our results provide phylogenetic constraints to the emergence of nitrogen fixation and suggest that nitrogenases likely emerged from maturase-like predecessors. Though the precise functional role of such a predecessor protein remains speculative, our results highlight evolutionary contingency as a significant factor shaping the evolution of a biogeochemically essential enzyme.

INTRODUCTION

The modern biosphere is shaped by a variety of essential and ancient enzymes that have co-evolved with the Earth environment for billions of years. Though general mechanisms for the gain of novel enzymatic functions have been explored (Ohno 1970; Gerlt and Babbitt 2001; Noda-Garcia et al. 2018; Copley 2021), the co-evolutionary steps toward the origins of many specific, key enzymes during Earth's early history remain unresolved. An unavoidable task in addressing this challenge is constraining the ancestral functions of early-evolved enzymes (Benner 2017; Hochberg and Thornton 2017, Kacar et al. 2017).

Biological nitrogen fixation is a notable example of a critical metabolic process with ancient and enigmatic origins. All life requires fixed, or bioavailable, nitrogen. For much of Earth history, this biologically vital element has primarily been acquired by organisms via the activities of nitrogenase metalloenzymes, early evolved and conserved catalysts that uniquely reduce dinitrogen (N_2) to ammonia (NH_3) (Hoffman et al. 2014; Einsle and Rees 2020). The evolution of nitrogenases has constrained the long-term productivity of the biosphere and has itself been shaped by the co-evolving biogeochemistry of Earth's environment (Falkowski 1997; Stüeken et al. 2016; Luo et al. 2018; Allen et al. 2019; Mus et al. 2019; Garcia et al. 2020). Nitrogenases represent the only known biomolecular solution for the reduction of N_2, a remarkable innovation given that the N≡N bond is one of the most inert in nature. The answer to how biology converged on this solution billions of years ago remains elusive (Boyd and Peters 2013; Mus et al. 2019; Garcia et al. 2020).

Insights into the origins of biological nitrogen fixation can be gained by reconstructing the protein ancestors of nitrogenases and their close homologs. Nitrogenases do not operate alone, but instead exist within a larger protein network required for their assembly and function. The closest homologs to nitrogenases are themselves a key step in nitrogenase assembly, serving as maturases (also referred to as assembly scaffold proteins) for the final steps in cofactor biosynthesis (Fay et al. 2016; Buren et al. 2020). These maturases are considered necessary for nitrogen-fixing organisms to modify a complex metal cluster precursor that, when matured, serves

as the nitrogenase active site for N_2 reduction. Though not themselves known to reduce N_2, these maturases reduce a variety of other substrates *in vitro*, including C_2H_2, CO, and CN^-, that also serve as alternative, non-physiological substrates of nitrogenases (Hu et al. 2010; Fay et al. 2016; Seefeldt et al. 2020). These findings establish maturases as catalytically similar to nitrogenases with the exception of their inability to reduce N_2. The divergent protein features between nitrogenases and maturases must therefore account for this difference in N_2-reduction capability.

To investigate the origins of biological nitrogen fixation, we reconstructed the evolutionary history of nitrogenases and maturases. By comparing patterns of sequence conservation between nitrogenases and maturases, we identified divergent residues that might account for their functional differentiation—namely, their ability or inability to reduce N_2, respectively. These sequence features were then leveraged to infer the N_2-reduction capability of reconstructed ancestral proteins and phylogenetically map the origins of biological nitrogen fixation within the nitrogenase family evolutionary history. Our findings support an origin of canonical nitrogenases from predecessor proteins that were unlikely to have been capable of N_2 reduction and more closely resemble extant maturases—proteins that are today only ancillary in biological nitrogen fixation. Nitrogenases may therefore represent a case in molecular evolution where a pre-existing but already complex protein architecture, adapted to an alternative role, shaped the origins of one of the most consequential biomolecular innovations in Earth history.

RESULTS AND DISCUSSION

Uncharacterized homologs root canonical nitrogenases within maturase protein clades

We reconstructed the phylogenetic history of nitrogenase and maturase homologs to explore ancestral states for this protein family. There are three forms of nitrogenase—Nif, Vnf, and Anf—that each differ in the composition of their active-site iron-sulfur cluster ("M-cluster"; incorporating molybdenum, vanadium, or additional iron, respectively). The heterotetrameric ($\alpha_2\beta_2$) catalytic protein component of nitrogenase (Nif/Vnf/AnfDK) has its counterpart in a homologous maturase protein complex (Nif/VnfEN; the Anf nitrogenase system does not have dedicated maturase proteins) (fig. 1a). We compiled a comprehensive dataset including nitrogenase Nif/Vnf/AnfDK and maturase NifEN protein sequences (fig. 1b), as well as outgroup dark-operative protochlorophyllide oxidoreductase homologs that share the $\alpha_2\beta_2$ subunit arrangement (BchNB) (Fujita and Bauer 2000; Moser and Brocker 2011). Four maximum likelihood phylogenies were built to test the robustness of tree topology and downstream ancestral sequence inference to sequence sampling and alternate alignment methods: 1) 2,425 nitrogenase, maturase, and outgroup homologs aligned by MAFFT (Katoh and Standley 2013), Tree-1; 2) removal of "uncharacterized" nitrogenase and maturase homologs, Tree-2; 3) removal of β-subunit nitrogenase, maturase, and outgroup homologs, Tree-3; and 4) alignment with MUSCLE (Edgar 2004) instead of MAFFT (see discussion below; Tree-1 to Tree-4; table 1).

General shared features across the reconstructed phylogenies include clustering of α-subunit sequences sister to β-subunit sequences (except in Tree-3, which lacks β-subunit sequences), reproducing the $\alpha_2\beta_2$ structural distinction for nitrogenases and maturases (fig. 1c, supplementary fig. S1). These α- and β-subunit clades themselves each segregate into nitrogenase and maturase protein sequences. This topology is consistent with an initial gene duplication event that resulted in separate α- and β-subunits, followed by a secondary duplication event that resulted in functionally distinct nitrogenase and maturase proteins (Fani et al. 2000; Raymond et al. 2004; Boyd and Peters 2013). Within the nitrogenase clade, vanadium- and iron-nitrogenase sequences nest within molybdenum-nitrogenase clades, as has been previously observed (Raymond et al. 2004; Boyd et al. 2011; Garcia et al. 2020). By contrast, the phylogenetic clustering of maturase sequences associated with different metal-dependent forms of nitrogenases do not reproduce this nesting pattern. Vanadium-maturase sequences are each split into two groups: one forms a small clade with a relatively long branch that diverges prior to all other nitrogenase and maturase sequences, whereas another diverges relatively recently within nitrogenase clades associated with aerobic or facultative bacteria. This topology suggests that assembly proteins for the vanadium nitrogenase system originated independently at least twice, with one origin associated with an early divergence from ancestors of unknown function and another from a recent nitrogenase ancestor.

For the phylogenies that include them (i.e., Tree-1, Tree-3, Tree-4), several maturase-like homologs diverge prior to nitrogenases and root the latter within the maturase clade (fig. 1c, supplementary fig. S1). These homologs belong to “uncharacterized” bacterial and archaeal taxa that lack extensive experimental characterization regarding the metal dependence and N_2-reducing capability of their nitrogenase-like proteins (McGlynn et al. 2012; Garcia et al. 2020). We assign these uncharacterized maturase-like sequences as putative maturases (rather than nitrogenases) based on three lines of evidence. First, the genes that encode these homologs in uncharacterized taxa are located closely downstream of nitrogenase-like genes, as is frequently the case with *bona fide* maturases (fig. 1b). Second, certain uncharacterized taxa (including those that are missing a NifN-like maturase subunit gene) have been shown to fix nitrogen, evidencing a functioning nitrogenase and, by extension, maturase (Mehta and Baross 2006; Dekas et al. 2009; Chen et al. 2021). Third, these homologs conserve a Cys48 residue (numbering here and hereafter from aligned *Azotobacter vinelandii* nitrogenase NifD) present in most maturase proteins (with the exception of certain VnfE homologs) and considered important for binding the cluster precursor prior to maturation (fig. 2a) (Kaiser et al. 2011). At the same time, these sequences all lack the strictly conserved nitrogenase His442 residue that ligates the active-site M-cluster and is critical for N_2 reduction (Kim and Rees 1992; Lee et al. 1998; Li 2002; Jimenez-Vicente et al. 2018). Together, these observations suggest that the uncharacterized maturase-like homologs are unlikely to be functioning as nitrogenases and are more likely operating as canonical maturases.

Our alternate tree reconstructions demonstrate that the rooting of nitrogenases within maturase clades is primarily determined by the inclusion of uncharacterized homologs. Though the phylogenetic position of certain uncharacterized clades is ambiguous (e.g., one clade diverges immediately prior to nitrogenases in Tree-1, but prior to both nitrogenases and other maturases in Tree-3 and Tree-4), we do not observe rooting of maturases within nitrogenase sequences in any of these trees (fig. 1c, supplementary fig. S1). This consistency suggests that the observed rooting pattern is robust to the exclusion of the β-subunit protein subtree (Tree-3) and variation in

tested alignment methodology (Tree-4). By contrast, the exclusion of uncharacterized homologs from the sequence dataset results in reciprocal monophyly between α-subunit nitrogenase and maturase clades and nesting of β-subunit maturase sequences within nitrogenase sequences (Tree-2; supplementary fig. S1). The α-subunit topology is therefore more stable and less affected by the presence of uncharacterized homologs than the β-subunit topology, which is consistent with the comparatively low branch support values among β-subunit sequences in Tree-2.

The rooting of nitrogenases within the broader maturase protein clade might parsimoniously suggest that the functionality of the common ancestor of both protein groups more likely resembled that of extant maturases. These phylogenetic observations present a hypothesis that can be tested by evaluating sequence features of reconstructed ancestral proteins inferred to contribute to the functional divergence between nitrogenases and maturases.

Phylogenetically divergent protein sites between extant nitrogenases and maturases map to functionally important structural regions

We performed a comprehensive, comparative sequence-structure analysis to identify protein sites that likely contribute to the functional divergence (i.e., (in)ability to reduce N_2) between nitrogenases and maturases. Our goal was to subsequently leverage this analysis for identification of similar sites in reconstructed protein ancestors and phenotypic inference. We limited our analysis to α-subunit nitrogenase (NifD) and maturase (NifE) sequences in part due to the greater topological uncertainty within β-subunit subtrees across our phylogenetic reconstructions (fig. 1c, supplementary fig. S1). In addition, α-subunit sequences host the active-site M-cluster or cluster precursor. We therefore expected that functional differences between nitrogenases and maturases are more likely to be modulated by sequence- and structural-level differences between α-subunit proteins.

To predict functionally divergent protein sites, we first calculated the amino acid frequency distributions of each alignment column for nitrogenase and maturase sequences (Materials and Methods). The Jensen-Shannon (J-S) distance between the two protein groups for every alignment column was calculated, where larger distances indicate greater divergence for that site between nitrogenases and maturases. We estimated the expected distribution of site-wise J-S distances by randomly partitioning our protein sequences into 2 groups 10,000 times and calculating site-wise J-S distances from each random partition. The p-value for each site's J-S distance from the nitrogenase-maturase partition was calculated from the distribution of J-S distances for that site across random sequence partitions. We defined functionally divergent sites as those exceeding the 75th-percentile distance across all alignment columns, as well as having an FDR-corrected p-value <0.0001. This analysis was repeated for all four alignments used to build phylogenetic trees (table 1), yielding 117 (116 for the Tree-4 alignment) predicted sites (supplementary table S1).

Putative functionally divergent sites identified from the sequence alignments cluster with known functionally important structural regions of the nitrogenase subunit (fig. 2b). These include the M-cluster active site for N_2 reduction, the interface between nitrogenase NifD and NifK subunits, and the M-cluster insertion funnel (Hu et al. 2008). This observed correlation suggests that the set of putative functionally divergent sites is likely enriched for positions contributing to the functional

divergence between nitrogenases and maturases. Six sites are of the most divergent across all sequence alignments: 69, 189, 362, 383, 440, and 444 (supplementary table S1). Two additional sites, 442 and 445, are highly divergent in alignments for Tree-1, Tree-2, and Tree-3, but not for Tree-4, likely because Tree-4 was constructed by a different alignment method that would have impacted downstream distance calculations (table 1). All are in or proximal to the nitrogenase active site or M-cluster insertion funnel. Some have specific inferred or experimentally determined functional roles in nitrogenases, including as an M-cluster ligand (site 442) (Kim and Rees 1992), a “lock” to hold the M-cluster within the active site (site 444) (Hu et al. 2008; Solomon et al. 2020), and a “lid” at the cluster insertion funnel opening (site 362) (Hu et al. 2008). Finally, our set of divergent sites includes the maturase cluster precursor ligand, site 48 (Kaiser et al. 2011). The clustering of highly divergent sites with these protein regions highlights the differential interaction of nitrogenases or maturases with the M-cluster. Specifically, maturase function does not require the conservation of residues that permit the insertion and stabilization of the cluster within the active site (Hu et al. 2008), nor the fine-tuning of residues in the active site for catalysis.

In addition to divergent sites, we identified several conserved residues among nitrogenase and maturase sequences (fig. 2a). His83, Cys88, Cys154, Gly160, Gly185, Gly194, and Cys275 residues are conserved in both nitrogenases and maturases. These sites are likely essential to both groups and arose prior to their evolutionary divergence. 23 sites are conserved only in nitrogenases, compared to just three sites that are conserved only in maturases: Cys62, Gly246, and Gly455. However, these three residues are still present in most nitrogenase sequences. The greater number of uniquely conserved residues in nitrogenases relative to maturases may reflect the stronger selective constraint associated with N_2-reduction functionality.

Putative functionally divergent sites of oldest ancestors resemble extant maturases more than nitrogenases

With a set of predicted functionally divergent protein sites between extant α-subunit nitrogenase and maturase proteins, we used a probabilistic approach to compare features of ancestral proteins inferred from all reconstructed phylogenies (table 1, supplementary fig. S1). For divergent sites, J-S distances were calculated between ancestral amino acid posterior probability distributions and extant amino acid frequency distributions for either extant nitrogenase or maturase homologs. These distance scores were then normalized to yield a value between -1 and +1, here called the “D-score” (supplementary table S2). Positive D-scores indicate greater similarity to α-subunit nitrogenase NifD homologs and negative scores indicate greater similarity to NifE homologs (see Materials and Methods for additional details). D-scores were averaged across divergent sites for each ancestral node.

Ancestral protein sequences inferred from oldest phylogenetic nodes on average resemble extant maturase homologs more than nitrogenase homologs at predicted functionally divergent sites. Mean D-scores across divergent sites for nodes ancestral to all nitrogenase and maturase homologs (fig. 3, “nitrogenase/maturase last common ancestor 1”) range between -0.08 and 0.00. These values are low-magnitude relative to the full range of mean D-scores across all ancestral nodes (~0.2 to +0.2), which is expected due to the mixing of both nitrogenase- and maturase-like sequence features at oldest nodes. Nevertheless, the primarily negative mean D-score values at these nodes indicate greater sequence-level similarity to maturase Nif/VnfE sequences than to nitrogenase Nif/Vnf/AnfD sequences. The ambiguity of the Tree-3 node (D-score = 0.00) may

result from the removal of β-subunit sequences for this phylogeny that would otherwise form an outgroup to the α-subunit clade and constrain ancestral sequence composition at this node. At more recent ancestral nodes that exclude early diverged VnfE clades (which have long, less highly supported branches and, thus, ambiguous evolutionary context; "nitrogenase/maturase last common ancestor 2"), mean D-scores range between -0.10 and +0.02. The only tree that yields positive D-scores for these more recent nodes is that reconstructed from the MUSCLE alignment (Tree-4), indicating that this analysis is more sensitive to the tested alignment method than sequence sampling. Nodes associated with the "nitrogenase last common ancestor" yield positive D-scores for all tree, and therefore resemble nitrogenases more than maturases at predicted functionally divergent sites.

In addition to surveying mean ancestral similarity to extant nitrogenases or maturases across all putative functionally divergent sites, we also investigated site-wise D-score values, some of which have been experimentally determined to contribute to either nitrogenase or maturase function. These site-wise D-scores were assessed specifically along the phylogenetic transect between the last common ancestor of nitrogenases and the last common ancestor of all nitrogenase and maturases (supplementary table S2). Certain divergent sites that become "nitrogenase-like" early (i.e., prior to the nitrogenase ancestor) include site 195 (important for N_2 substrate binding (Kim et al. 1995)), site 444 (locks the M-cluster in the nitrogenase active site (Hu et al. 2008; Solomon et al. 2020)), and site 359 (helps form the cluster insertion funnel (Hu et al. 2006)). By contrast, site 48 (involved with L-cluster binding in maturases (Kaiser et al. 2011)), as well as sites 361 and 362 (involved with M-cluster insertion at the nitrogenase active site (Hu et al. 2008)), remain primarily "maturase-like" until the nitrogenase last common ancestor.

The site 442 histidine M-cluster ligand conserved across all extant canonical nitrogenase proteins is considered critical for nitrogenase activity (Kim and Rees 1992; Lee et al. 1998; Li 2002; Jimenez-Vicente et al. 2018) and, notably, remains maturase-like prior to the nitrogenase last common ancestor (mean D-score ≈ -0.40 across alternate alignments, compared to a minimum D-score of -0.76 for all divergent sites; supplementary table S2). Site 442 is only not predicted to be functionally divergent for the tree reconstructed by a MUSCLE alignment (Tree-4). This difference is likely due to an inferred homology, unique to the MUSCLE alignment, between the nitrogenase His442 residue and a frequently observed histidine residue in maturases. However, the other alignments, which instead infer homology with a neighboring arginine residue in many maturases (fig. 2a), is supported by studies which indicate that this arginine residue is structurally aligned to nitrogenase His442 (Kaiser et al. 2011). Thus, it is likely that the MUSCLE alignment is erroneous at this site. Due to its functional significance and conservation among extant nitrogenases, the appearance of a histidine residue at site 442 may have been critical for the origins of N_2 reduction.

Nitrogenases likely originated from a non-N_2-reducing maturase-like protein with possible biosynthetic or alternate catalytic roles

Our exploration of nitrogenase and maturase ancestry, coupled with the investigation of global sequence and structural features that account for their functional divergence, suggests that the last common ancestor of nitrogenase and maturase proteins likely resembled the latter and was unlikely to have functioned as a nitrogenase. This inference is also supported by the absence of

residue-level similarity at divergent sites that have empirically been shown to be critical for nitrogenase function. These inferences are robust to phylogenetic uncertainty stemming from both sequence sampling of early diverged uncharacterized lineages and alignment method, and incorporate statistical uncertainty associated with ancestral sequence inference.

A model that posits a maturase-like ancestry for nitrogenases deviates from existing hypotheses regarding their early evolution (fig. 4). Previous models are based on parsimonious interpretations of nitrogenase and maturase phylogenetic topology that is not observed in the trees reconstructed here. For instance, previous studies root the maturase clade within nitrogenase sequences, suggesting that the former evolved via gene duplication of nitrogenase ancestors (Boyd et al. 2011; Boyd and Peters 2013). However, with the inclusion of uncharacterized nitrogenase and maturase homologs, maturases instead root nitrogenases, and support a model wherein nitrogen fixation is instead a derived feature of a maturase-like ancestor. Our reconstruction of ancestral states within the nitrogenase and maturase phylogeny provides additional constraints on ancestral phenotypes within a maximum likelihood framework, extending beyond inferences drawn from phylogenetic topology alone. These results constrain the likely origin of nitrogen fixation to a relatively well-resolved lineage within the nitrogenase/maturase topology, rather than to a deeper history that bridges nitrogenases with more distantly related homologs (e.g., coenzyme F430 biosynthesis proteins (Mus et al. 2019)). Conservatively, our results suggest that phylogenetic inference of ancient nitrogen fixation can reliably extend only to within the canonical nitrogenase clade, and that inferences of earlier nitrogen fixation activity require further evidence. More definitive assessment of the N_2-reduction capability of the common ancestor of nitrogenases and maturases awaits experimental investigation, which can be directly achieved through the laboratory resurrection of ancestral proteins inferred in this study (Thornton 2004; Benner 2017; Garcia and Kacar 2019).

Our results suggest that ancestral maturase-like proteins provided the molecular architecture for the origin of nitrogen fixation. However, the precise functional role of this ancestor is not clear. Candidate phenotypic attributes, shared between extant nitrogenases and maturases, may shed light on how biological N_2 reduction evolved. For example, both are the only proteins known to bind the nitrogenase active-site M-cluster (Fay et al. 2016). In addition, both nitrogenases and maturases can reduce several non-physiological, alternative substrates including C_2H_2, CO, and CN^- (Fay et al. 2016; Seefeldt et al. 2020). It has previously been argued that the ability of nitrogenases to reduce alternative substrates may simply be the byproduct of overcoming the significant activation barrier required for N_2 reduction (Boyd and Peters 2013). However, combined with the evidence presented here for a maturase-like ancestry, we suggest that the most parsimonious scenario is an ancestral protein capable of reducing the shared substrates of extant nitrogenases and maturases at an M-cluster-like active site (perhaps in a role as detoxyases, as has previously been proposed (Silver and Postgate 1973); fig. 4). This scenario provides a stepwise path for the evolution of nitrogenases from ancestral proteins capable of catalyzing less ATP-intensive reactions (Hu et al. 2010), and requires only residue-level tuning of an already complex peptide environment to achieve the earliest whiffs of N_2 reduction. Our reconstruction of ancestral residues at predicted functionally divergent sites provides possible mutational trajectories toward the evolution of nitrogenases.

The origin of nitrogenases from proteins involved with cofactor biosynthesis such as the maturases studied here is also not unexpected given the prevalence of biosynthetic proteins

associated with the broader family of nitrogenase-like homologs. These include chlorophyll biosynthesis proteins used as phylogenetic outgroups in our study (Fujita and Bauer 2000; Moser and Brocker 2011), as well as coenzyme F430 biosynthesis proteins that are conserved among methanogens (Staples et al. 2007; Zheng et al. 2016). In addition, more distantly related and poorly studied nitrogenase-like homologs may have roles in assembly of hydrogenase metalloclusters and metal transport (reviewed in Ghebreamlak and Mansoorabadi (2020)). It is not clear whether a maturase-like nitrogenase ancestor may have had a central function in cofactor biosynthesis as its descendants and several related homologs do today. Other distantly related nitrogenase-like homologs have putative catalytic roles, including a recently reported homolog suggested to participate in a methionine salvage pathway that forms ethylene (North et al. 2020). It is possible that maturase-like predecessors were promiscuous, a suggested general feature of early-evolved proteins (Copley 2015, De Tarafder et al. 2021), and capable of both providing a scaffold for cluster maturation as well as catalysis using the same matured cluster. Gene duplication and divergence may have subsequently specialized maturases to only function as a scaffold and evolve residues that permit the release of the matured cluster. In parallel, nitrogenases may have specialized to function in a solely catalytic role.

The possible capability of ancestral maturase-like proteins to bind an M-cluster-like cofactor has important implications for the coevolutionary trajectory of nitrogenases and the biogeochemical environment. A defining role of extant maturases in the molybdenum-dependent nitrogenase system is to provide a scaffold for the incorporation of molybdenum into the active-site M-cluster. Thus, the origin of maturase proteins has previously been suggested to coincide with the origin of molybdenum dependence in nitrogenases (Boyd et al. 2011). However, if maturase-like proteins predate nitrogenases, the molybdenum-containing M-cluster itself may predate nitrogenases as well (fig. 4). Inferences for the age of nitrogenases extend to more than 3 billion years ago (Stueken et al. 2015; Parsons et al. 2020) when molybdenum in Earth's oceans was likely exceedingly scarce (Anbar 2008). A molybdenum-incorporating maturase-like protein existing prior to 3 billion years ago would suggest that the bulk geochemistry of the early Earth environment may not necessarily have provided strict constraints on enzyme evolution (Garcia et al. 2020), and would be in agreement with geochemical evidence of early molybdenum-dependent nitrogenase activity (Stueken et al. 2015). Rather, localized environments may have provided sufficient molybdenum for the function of nitrogenases and their predecessors (Stueken et al. 2015), or molybdenum may have simply been selected despite its scarcity due to its invaluable chemical features. Another possibility is that earliest maturases-like ancestors did not incorporate molybdenum, but rather the iron-only cluster precursor that is matured to the M-cluster today. The extant precursor-bound maturase protein is in fact capable of substrate reduction *in vitro* (Hu et al. 2010), presenting a possible ancestral state in the absence of abundant molybdenum.

The role of contingency and subsumed complexity in nitrogenase evolution

The model of nitrogenase origins that we propose exemplifies a case for molecular evolution in which a novel and consequential metabolic function was built off a pre-existing, complex molecular architecture. An open question is whether the pre-existing complexity and functional role of a maturase-like protein was needed to evolve an enzyme capable of reducing N_2. The role of evolutionary contingency in shaping biological diversity has long been examined (Gould 1989; Vermeij 2006; Blount et al. 2018), particularly to envision future evolutionary scenarios or alternate

trajectories on other worlds characterized by distinct environmental parameters (Kaçar and Gaucher 2012). Regarding the evolution of biological nitrogen fixation in particular, it has been argued that necessity—i.e., the need for bioavailable sources of nitrogen—or environmental geochemistry likely controlled the timing of early nitrogenase evolution and diversification on Earth (Navarro-Gonzalez et al. 2001; Anbar and Knoll 2002; Mus et al. 2018). However, another possibility is that the origin of biological nitrogen fixation required the subsumed complexity of a protein predecessor (Adam et al. 2018), which was initially positively selected for an entirely different metabolic role. The origin of nitrogenases thus may not have occurred (or may have been significantly delayed) without a suitable protein on which to build, despite the scarcity of bioavailable nitrogen, an abundance of possible metal cofactors, or both. Testing these possibilities would require experimentally replaying the evolutionary path that led to the origin of biological nitrogen fixation.

There is no evidence that Terran biology evolved nitrogen fixation more than once. Whether this is the product of very exceptional circumstances of origination or of a survivorship bias so pronounced that there is only one remaining functional example supported across all of biology, nitrogenases are therefore on par with other singular molecular-level innovations such as the ribosome (Fox 2010) and oxygenic photosynthesis (Blankenship and Hartman 1998). The existing variation among different metal-dependent forms of nitrogenase enzymes do not constitute truly independent evolutionary experiments, but variations on a theme that was determined and uniquely constrained by the common ancestral form. Even within this narrow range of constraints, the degree to which contingent amino-acid substitutions shaped the diversification of nitrogenase metal preference and specificity remains unknown.

Identifying the singular circumstances that left Earth with a single common ancestor for all nitrogenase function may be critical for understanding the pervasiveness of nitrogen fixation as a universal biological capability. This is particularly important for assessing whether metal availability significantly guides protein evolution or whether other internal biophysical constraints lead to background-dependent, epistatic interactions. Efforts to generate artificial nitrogenases and nitrogenase metalloclusters (Tanifuji et al. 2015; Sickerman et al. 2017) may expand the suite of molecular structures capable of reducing N_2, but biotic experiments integrating gene regulatory and protein-protein interaction constraints are needed to test different macroevolutionary hypotheses of nitrogenase emergence. A survey of such functional constraints on nitrogenase and maturase predecessors could reveal the sequence of biomolecular functions conducive for the evolution of nitrogen fixation, which could then be integrated into a more comprehensive accounting of internal selective forces, geochemical features and planetary environments that can host similar evolutionary pathways (Kacar et al. 2020).

Perhaps most intriguingly, our results suggest that nitrogen fixation may have emerged from natural selection acting on a maturase-like protein whose ancestral function was largely decoupled from extracellular conditions frequently implicated as drivers for the origins of nitrogen fixation. This scenario casts the origin of nitrogen fixation as an act of extreme contingency bordering on happenstance, betraying its utility as one of the most evolutionarily significant and biologically limiting metabolic pathways on Earth. If borne out by further study or found to be a recurring pattern for other critical molecular innovations, enzyme origins largely decoupled from putative environmental drivers may severely compromise the soundness of systems-focused hypotheses that tie organismal or ecological need to bulk geochemical substrate or cofactor

availabilities. The paleobiology of molecular innovations would require disciplinary approaches and conceptual foundations quite distinct from the study of their more recent evolution.

MATERIALS AND METHODS

Phylogenetic reconstruction and ancestral sequence inference

Protein homologs were identified from the National Center for Biotechnology Information non-redundant protein database by BLASTp (Camacho et al. 2009) with an expect value cutoff of <1e-5 (accessed January 2020). Query sequences from *Azotobacter vinelandii* (NifD: WP_01270336, NifK: WP_012698833; NifE: WP_012698838, NifN: WP_012698839) were used for nitrogenase and maturase homolog identification, and sequences from *Synechococcus elongatus* (BchN: WP_126148028, BchB: WP_126147769) for outgroup dark-operative protochlorophyllide oxidoreductase homolog identification. Sequences were aligned using MAFFT v7.450 (Katoh and Standley 2013) to build a preliminary phylogeny with FastTree v2.1.11 (Price et al. 2010). Sequences in overrepresented clades were pruned to obtain a roughly equal number of nitrogenase and assembly protein versus outgroup sequences, so as not to bias subsequent ancestral sequence inference. The final dataset was realigned by MAFFT and used as input for phylogenetic reconstruction by RAxML v8.2.11 using the LG+G+F evolutionary model (Stamatakis 2014). The tree was further optimized by applying nearest-neighbor-interchanges before calculation of SH-like branch support values (Guindon et al. 2010), resulting in the final phylogeny, Tree-1. Additional trees were generated by altering sequence sampling or alignment with MUSCLE v3.8.425 (Edgar 2004) instead of MAFFT (see Results and Discussion, table 1).

Ancestral sequences were inferred by maximum likelihood marginal reconstruction in PAML v4.9j (Yang 2007) using the same evolutionary model parameters described above for RAxML. Sequence gaps were reconstructed in PAML using the binary encoding approach described in Aadland et al. (2019). Briefly, the protein sequence alignment was recoded as a ‘presence-absence’ alignment matrix, and the posterior-probability of the presence (amino-acid residue) or absence (gap) state at each position in each ancestral sequence was calculated using maximum-likelihood reconstruction, assuming a binary character model with state frequencies inferred by maximum likelihood.

Prediction of functionally divergent protein sites between nitrogenases and assembly proteins

For each position in the sequence alignment, we calculated the Jensen-Shannon (J-S) distance between the amino-acid frequency distribution estimated from extant nitrogenase Nif/Vnf/AnfD sequences in the alignment and that estimated from extant maturase Nif/VnfE sequences, with nitrogenase and maturase sequences being defined based on monophyly in the tree topology and gene location within the *nif*, *vnf*, or *anf* loci (see Results and Discussion, fig. 1b). Briefly, the J-S distance is calculated as the average Kullback-Leibler divergence, or “relative entropy,” which estimates the loss of information when one frequency distribution is used to represent another. Intuitively, the site-wise J-S distance between nitrogenase and maturase amino-acid frequency

distributions describes how dissimilar the distribution of extant amino-acids is between nitrogenase and maturase sequences for each alignment site.

We estimated the expected distribution of site-wise J-S distances, given our sequence alignment, by randomly partitioning the alignment into two sequence groups of sizes equivalent to the sizes of our actual nitrogenase and maturase groups, respectively, and calculating site-wise J-S distances between these randomly partitioned groups. We performed 10,000 random partitions and site-wise J-S distance calculations. For each alignment column *i*, we calculated the probability of observing J-S(nitrogenase, maturase)$_i$, given the distribution of J-S-distances at column *i* in the randomly partitioned dataset (i.e., *p*-value). We enriched for highly divergent positions using an FDR-corrected *p*-value cutoff of 0.0001. We additionally excluded any sites with J-S distances in the lower 75th percentile of the J-S distance distribution across all alignment positions.

Probabilistic assessment of nitrogenase-like ancestral sequence features

Nitrogenase-like ancestral sequence features were assessed by incorporating the statistical uncertainty of ancestral sequence inference into comparisons between those of extant nitrogenases and maturases. For putative functionally divergent protein sites identified as described above, J-S distances were calculated between the ancestral amino acid posterior probability distributions and the extant amino acid frequency distributions across either extant nitrogenase (Nif/Vnf/AnfD) or maturase (Nif/VnfE) clades. These distance values were then normalized to yield a value between -1 and +1 indicating the relative similarity of an ancestral protein site to a homologous extant nitrogenase site, here called the “D-score” (i.e., similarity to the nitrogenase D-subunit):

$$\text{D-score} = (\text{dist}_E - \text{dist}_D) / (\text{dist}_E + \text{dist}_D)$$

where dist_D is the J-S distance between ancestral and maturase E-subunit sites, and dist_E is the J-S distance between ancestral and nitrogenase D-subunit sites. D-scores were analyzed on a site-wise basis as well as averaged across the length of each ancestral sequence for all constructed phylogenies.

ACKNOWLEDGEMENTS

This work was supported by the National Aeronautics and Space Administration Postdoctoral Program to A.K. Garcia., the National Science Foundation Division of Molecular and Cellular Biosciences (1817942 to B. Kolaczkowski), the National Science Foundation Emerging Frontiers Program (1724090 to B. Kaçar), the National Aeronautics and Space Administration Early Career Faculty Award (80NSSC19K1617 to B. Kaçar), and the Metal Utilization and Selection across Eons (MUSE) Interdisciplinary Consortium for Astrobiology Research, sponsored by the National Aeronautics and Space Administration Science Mission Directorate (19-ICAR19_2-0007 to B. Kaçar). We thank Eric Boyd and Jennifer Glass for the helpful discussions.

517 **TABLES**

518

519 **Table 1.** Nitrogenase and assembly-protein phylogenies built in this study.

520

Tree	Alignment method	Sequence Dataset
Tree-1	MAFFT	Nif/Vnf/AnfDK, Nif/VnfEN, BchNB
Tree-2	MAFFT	Tree-1 dataset without uncharacterized homologs
Tree-3	MAFFT	Tree-1 dataset without β-subunit sequences (Nif/Vnf/AnfK, Nif/VnfN, BchB)
Tree-4	MUSCLE	Same as Tree-1 dataset

521

522

523

FIGURES AND FIGURE LEGENDS

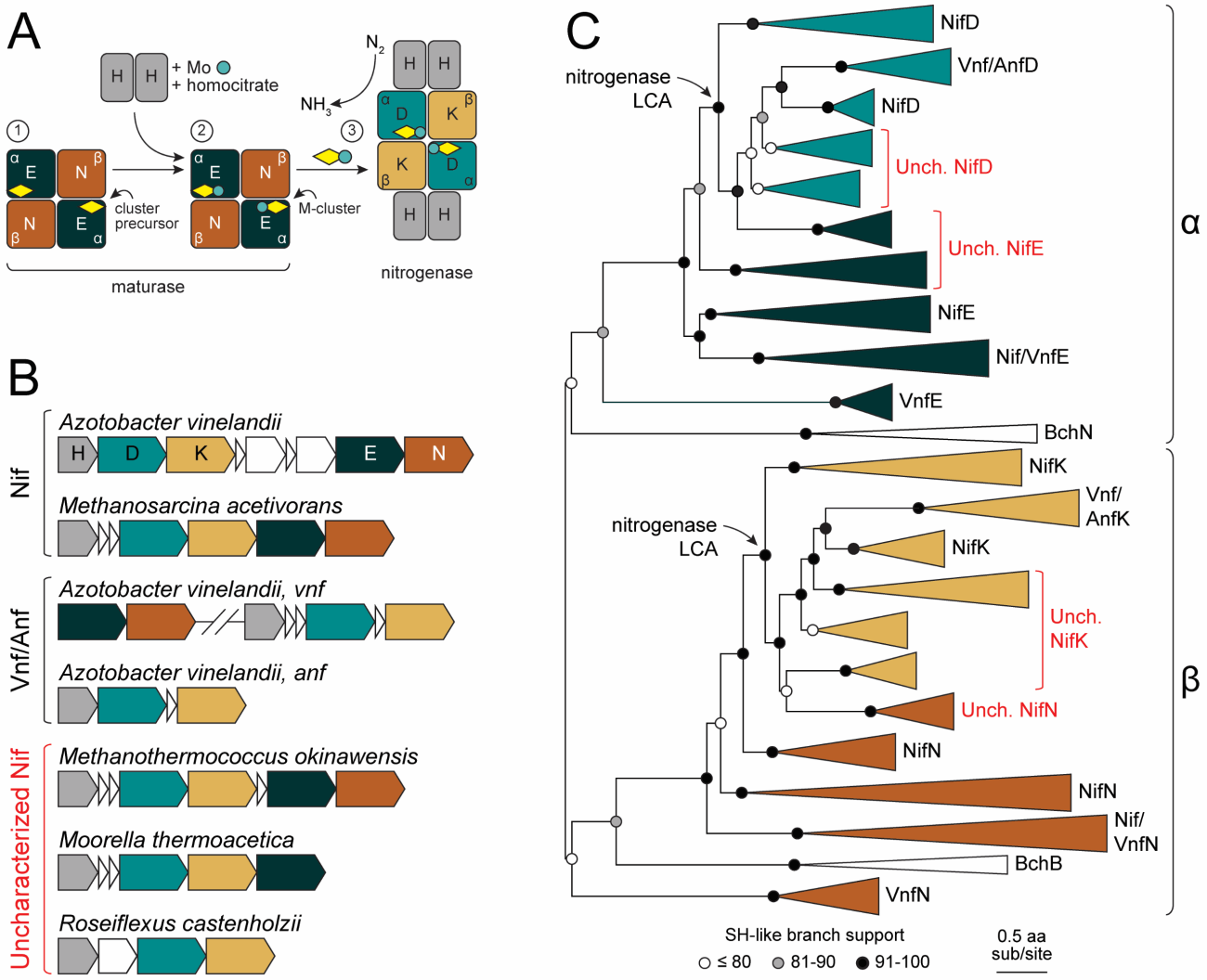
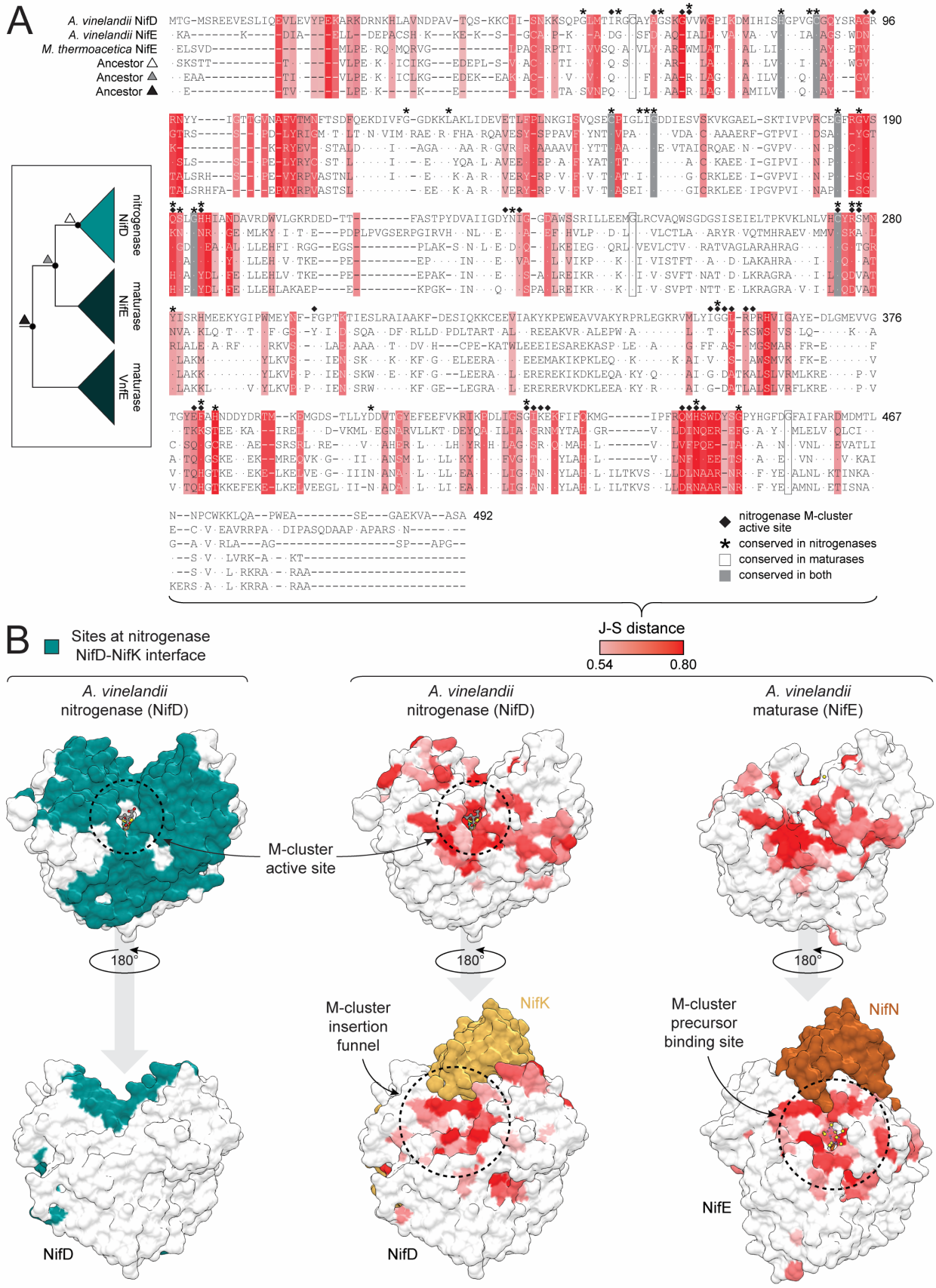


Fig. 1. Nitrogenase and maturase functionality, genetic organization, and evolutionary history. (*A*) Simplified biosynthetic pathway for nitrogenase active-site cluster maturation and incorporation (molybdenum-dependent nitrogenase system shown). Pathway steps are indicated by circled numbers. Step 1). Maturase NifEN proteins (NifE, dark green; NifN, brown), which are $\alpha_2\beta_2$ heterotetramer homologs to nitrogenase NifDK proteins (NifD, teal; NifK, yellow), are scaffolds for maturation of nitrogenase iron-sulfur cluster precursors (yellow diamond). Step 2) Cluster precursors are matured to M-clusters (yellow diamond with teal circle) by incorporation of molybdenum and homocitrate, delivered to the maturase complex by NifH proteins. Step 3) Mature M-clusters are incorporated into the nitrogenase complex, where they serve as the active sites for N_2 reduction to NH_3. During enzyme turnover, NifH proteins transiently interface with the nitrogenase NifDK complex to deliver electrons to the M-cluster active site. (*B*) Representative gene locus structures for molybdenum-dependent (Nif), vanadium-dependent (Vnf), and iron-dependent (Anf) nitrogenase systems, as well as for uncharacterized nitrogenase homologs (see text for discussion). Gene and intergenic region lengths are approximate. Hash marks indicate

543 significant distance between represented genes. (*C*) Maximum likelihood phylogenetic tree built
544 from nitrogenase and maturase protein sequences. α- or β-subunit designations for protein
545 sequences are indicated on the right. Homologs from uncharacterized taxa are highlighted in red.
546 Clade widths are not to scale. “LCA”: Last common ancestor.
547


A
A. vinelandii NifD
A. vinelandii NifE
M. thermoacetica NifE
Ancestor
Ancestor
Ancestor
nitrogenase NifD
maturase NifE
maturase VnfE
nitrogenase M-cluster active site
conserved in nitrogenases
conserved in maturases
conserved in both
B
Sites at nitrogenase NifD-NifK interface
J-S distance
0.54
0.80
A. vinelandii nitrogenase (NifD)
A. vinelandii nitrogenase (NifD)
A. vinelandii maturase (NifE)
M-cluster active site
180°
180°
180°
M-cluster insertion funnel
NifK
M-cluster precursor binding site
NifN
NifD
NifD
NifE

549 **Fig. 2.** Structure and sequence maps of putative functionally divergent protein sites between
550 nitrogenases and maturases. (*A*-*B*) Putative functionally divergent sites are defined as those
551 above the 75th percentile J-S distance between nitrogenase NifD and maturase NifE proteins (see
552 text for details). The J-S distance scale applies to both (*A*) and (*B*). (*A*) Maximum likelihood
553 ancestors (inferred from Tree-1; fig. 1c) aligned to representative extant nitrogenase (NifD) and
554 maturase (NifE) sequences. *Moorella thermoacetica (M. thermoacetica*) NifE is an
555 uncharacterized, putative maturase sequence. Ancestor triangle symbols correspond to labeled
556 nodes in the simplified phylogeny (*left*) and match those in fig. 3. Nitrogenase M-cluster active-
557 site residues are defined as those within 5 Å of the M-cluster. Dots within the alignment indicate
558 residue identity to *Azotobacter vinelandii* (*A. vinelandii*) NifD. Site numbering based on *A.*
559 *vinelandii* NifD. (*B*) Divergent sites mapped to aligned nitrogenase (*center*; *A. vinelandii* NifD, PDB
560 3U7Q) and maturase (*right*; *A. vinelandii* NifE, PDB 3PDI) subunit structures. Protein sites at the
561 nitrogenase NifD-NifK interface (*left*) are defined as those within 5 Å of the NifK subunit.
562
563
564

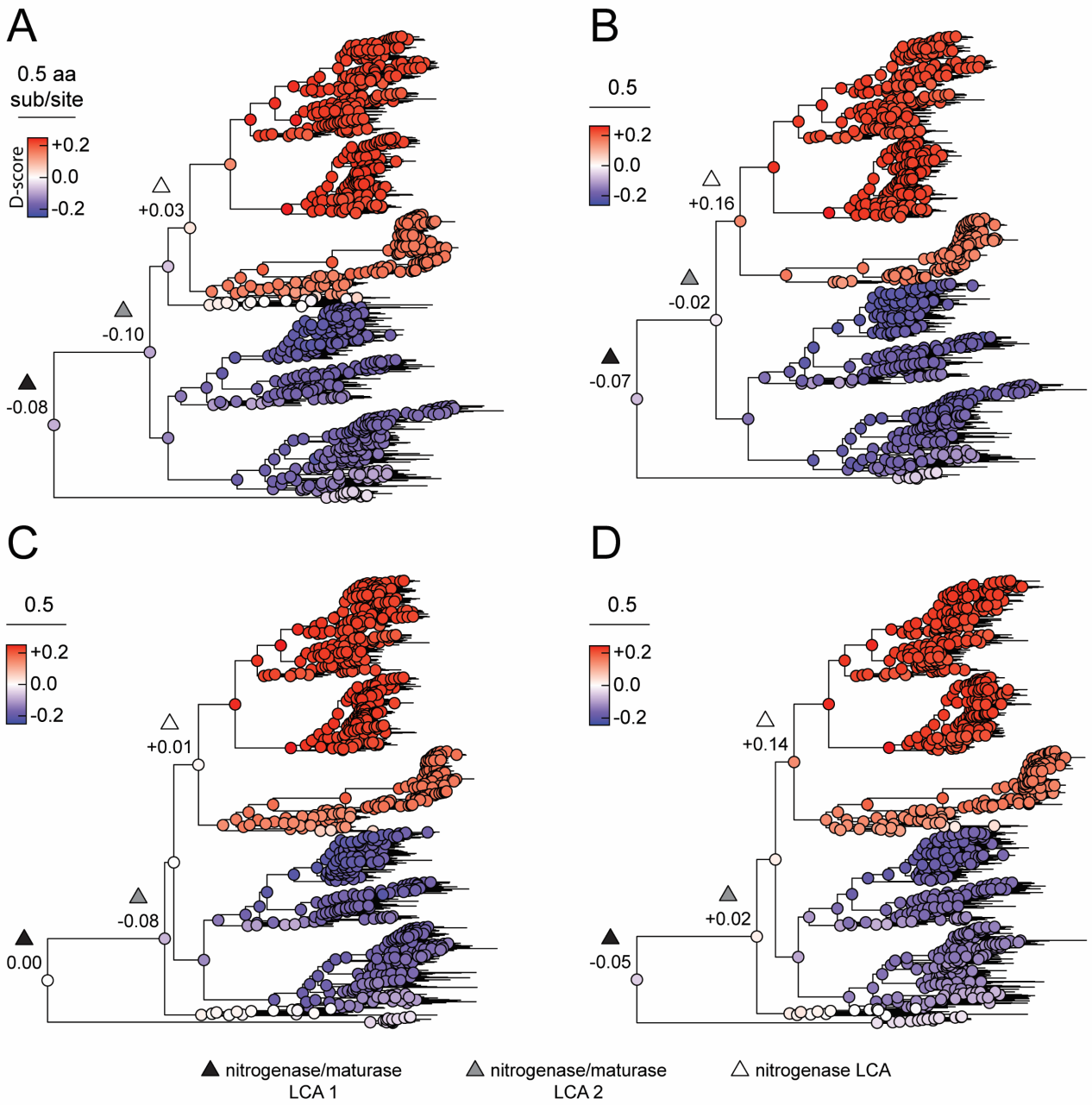


Fig. 3. Ancestral sequence similarity to extant nitrogenases or assembly proteins, mapped across four alternate phylogenies: (*A*) Tree-1, (*B*) Tree-2, (*C*) Tree-3, and (*D*) Tree-4. Similarity is expressed as the "D-score" parameter, where a positive D-score (red) indicates greater similarity to extant nitrogenases, and a negative D-score (blue) indicates greater similarity to extant maturases (see Materials and Methods). D-scores were averaged across putative functionally divergent sites for each ancestral node. Nodes are labeled nitrogenase/assembly protein last common ancestor ("LCA") 1—including early diverged VnfE homologs (black triangle; see text for discussion); nitrogenase /assembly protein LCA 2—excluding early diverged VnfE homologs (grey triangle); and nitrogenase LCA (white triangle), along with their mean D-scores.

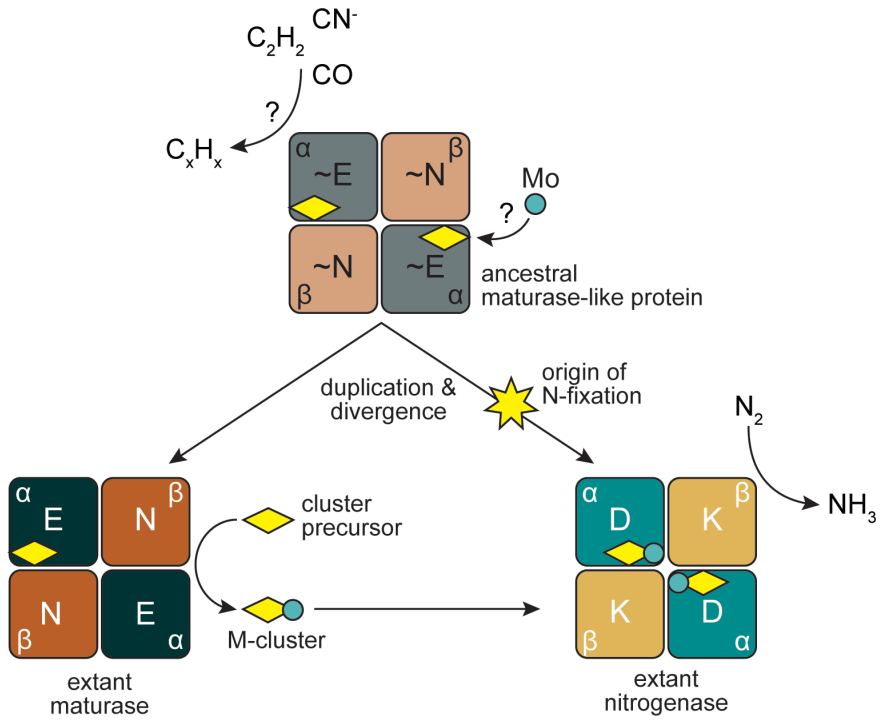


Fig. 4. Model for the origins and functional divergence of maturase and nitrogenase proteins. An ancestral maturase-like protein (~NifEN, grey and light brown), incapable of reducing N_2, may have otherwise reduced various carbon-containing substrates and/or played a role in cluster (yellow diamond) biosynthesis. The ancestor may have been capable of incorporating molybdenum (teal circle) into the cluster. Duplication of the encoding ancestral genes and functional divergence would then have yielded canonical maturase (NifEN, dark green and brown) and nitrogenase (NifDK, teal and yellow) proteins. Maturases would have specialized to provide a scaffold for the maturation of the nitrogenase cluster precursor (yellow diamond) to the nitrogenase active-site M-cluster (yellow diamond with teal circle). In parallel, tuning of the ancestral peptide environment along a divergent lineage would have spurred the origin of N_2 reduction and specialization of nitrogenases for a solely catalytic role in nitrogen fixation. Protein components of the $\alpha_2\beta_2$ heterotetrameric nitrogenase and maturase structures are labeled.

REFERENCES

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

Adam ZR, Fahrenbach AC, Kacar B, Aono M. 2018. Prebiotic Geochemical Automata at the Intersection of Radiolytic Chemistry, Physical Complexity, and Systems Biology. *Complexity*. 2018:1-21. doi: 10.1155/2018/9376183.

Allen JF, Thake B, Martin WF. 2019. Nitrogenase Inhibition Limited Oxygenation of Earth's Proterozoic Atmosphere. *Trends Plant Sci*. 24:1022-1031. doi: 10.1016/j.tplants.2019.07.007.

Anbar AD. 2008. Oceans. Elements and evolution. *Science*. 322:1481-1483. doi: 10.1126/science.1163100.

Anbar AD, Knoll AH. 2002. Proterozoic ocean chemistry and evolution: a bioinorganic bridge? *Science*. 297:1137-1142. doi: 10.1126/science.1069651.

Benner S. 2017. Uniting natural history with the molecular sciences. The ultimate multidisciplinarity. *Acc Chem Res*. 50:498-502. doi: 10.1021/acs.accounts.6b00496.

Blankenship RE, Hartman H. 1998. The origin and evolution of oxygenic photosynthesis. *Trends Biochem Sci*. 23:94-97. doi: 10.1016/S0968-0004(98)01186-4.

Blount ZD, Lenski RE, Losos JB. 2018. Contingency and determinism in evolution: Replaying life’s tape. *Science*. 362:eaam5979. doi: 10.1126/science.aam5979.

Boyd ES, Anbar AD, Miller S, Hamilton TL, Lavin M, Peters JW. 2011. A late methanogen origin for molybdenum-dependent nitrogenase. *Geobiology*. 9:221-232. doi: 10.1111/j.1472-4669.2011.00278.x.

Boyd ES, Peters JW. 2013. New insights into the evolutionary history of biological nitrogen fixation. *Front Microbiol*. 4:201. doi: 10.3389/fmicb.2013.00201.

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

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

Chen Y, Nishihara A, Haruta S. 2021. Nitrogen-fixing Ability and Nitrogen Fixation-related Genes of Thermophilic Fermentative Bacteria in the Genus *Caldicellulosiruptor*. *Microbes Environ*. 36:ME21018. doi: 10.1264/jsme2.ME21018.

Copley SD. 2015. An evolutionary biochemist's perspective on promiscuity. *Trends Biochem Sci*. 40:72-78. doi: 10.1016/j.tibs.2014.12.004.

Copley SD. 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.

630 De Tarafder A, Parajuli NP, Majumdar S, Kaçar B, Sanyal S. 2021. Kinetic Analysis Suggests
631 Evolution of Ribosome Specificity in Modern Elongation Factor-Tus from “Generalist”
632 Ancestors. *Mol Bio Evol*. doi: 10.1093/molbev/msab114
633
634 Dekas AE, Poretsky RS, Orphan VJ. 2009. Deep-sea archaea fix and share nitrogen in
635 methane-consuming microbial consortia. *Science*. 326:422-426. doi: 10.1126/science.1178223.

636 Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high
637 throughput. *Nucleic Acids Res*. 32:1792-1797. doi: 10.1093/nar/gkh340.

638 Einsle O, Rees DC. 2020. Structural Enzymology of Nitrogenase Enzymes. *Chem Rev*.
639 120:4969-5004. doi: 10.1021/acs.chemrev.0c00067.

640 Falkowski PG. 1997. Evolution of the nitrogen cycle and its influence on the biological
641 sequestration of CO_2 in the ocean. *Nature*. 387:272-275. doi: 10.1038/387272a0.

642 Fani R, Gallo R, Lio P. 2000. Molecular evolution of nitrogen fixation: the evolutionary history of
643 the *nifD*, *nifK*, *nifE*, and *nifN* genes. *J Mol Evol*. 51:1-11. doi: 10.1007/s002390010061.

644 Fay AW, Blank MA, Rebelein JG, Lee CC, Ribbe MW, Hedman B, Hodgson KO, Hu Y. 2016.
645 Assembly scaffold NifEN: A structural and functional homolog of the nitrogenase catalytic
646 component. *Proc Natl Acad Sci USA*. 113:9504-9508. doi: 10.1073/pnas.1609574113.

647 Fox GE. 2010. Origin and Evolution of the Ribosome. *CSH Perspect Biol*. 2:a003483-a003483.
648 doi: 10.1101/cshperspect.a003483.

649 Fujita Y, Bauer CE. 2000. Reconstitution of Light-independent Protochlorophyllide Reductase
650 from Purified BchI and BchN-BchB Subunits. *J Biol Chem*. 275:23583-23588. doi:
651 10.1074/jbc.M002904200.

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

654 Garcia AK, McShea H, Kolaczkowski B, Kacar B. 2020. Reconstructing the evolutionary history
655 of nitrogenases: Evidence for ancestral molybdenum-cofactor utilization. *Geobiology*. 18:394-
656 411. doi: 10.1111/gbi.12381.

657 Gerlt JA, Babbitt PC. 2001. Divergent Evolution of Enzymatic Function: Mechanistically Diverse
658 Superfamilies and Functionally Distinct Suprafamilies. *Annu Rev Biochem*. 70:209-246. doi:
659 10.1146/annurev.biochem.70.1.209.

660 Ghebreamlak SM, Mansoorabadi SO. 2020. Divergent Members of the Nitrogenase
661 Superfamily: Tetrapyrrole Biosynthesis and Beyond. *ChemBioChem*. 21:1723-1728. doi:
662 10.1002/cbic.201900782.

663 Gould SJ. 1989. Wonderful Life: The Burgess Shale and the Nature of History. New York, NY:
664 W. W. Norton & Company, Inc.

665 Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms
666 and methods to estimate maximum-likelihood phylogenies: assessing the performance of
667 PhyML 3.0. *Syst Biol*. 59:307-321. doi: 10.1093/sysbio/syq010.

668 Hochberg GKA, Thornton JW. 2017. Reconstructing ancient proteins to understand the causes
669 of structure and function. *Annu Rev Biophys*. 46:247-269. doi: 10.1146/annurev-biophys-
670 070816-033631.

671 Hoffman BM, Lukoyanov D, Yang ZY, Dean DR, Seefeldt LC. 2014. Mechanism of nitrogen
672 fixation by nitrogenase: the next stage. *Chem Rev*. 114:4041-4062. doi: 10.1021/cr400641x.

673 Hu Y, Fay AW, Lee CC, Wiig JA, Ribbe MW. 2010. Dual functions of NifEN: insights into the
674 evolution and mechanism of nitrogenase. *Dalton T*. 39. doi: 10.1039/b922555b.

675 Hu Y, Fay AW, Lee CC, Yoshizawa J, Ribbe MW. 2008. Assembly of Nitrogenase MoFe
676 Protein. *Biochemistry*. 47:3973-3981. doi: 10.1021/bi7025003.

677 Hu Y, Fay AW, Schmid B, Makar B, Ribbe MW. 2006. Molecular Insights into Nitrogenase
678 FeMoco Insertion. *J Biol Chem*. 281:30534-30541. doi: 10.1074/jbc.M605527200.

679 Jimenez-Vicente E, Yang ZY, Ray WK, Echavarri-Erasun C, Cash VL, Rubio LM, Seefeldt LC,
680 Dean DR. 2018. Sequential and differential interaction of assembly factors during nitrogenase
681 MoFe protein maturation. *J Biol Chem*. 293:9812-9823. doi: 10.1074/jbc.RA118.002994.

682 Kacar B, Hanson-Smith V, Adam ZR, Boekelheide N. 2017. Constraining the timing of the Great
683 Oxidation Event within the Rubisco phylogenetic tree. Geobiology. 155:628-40.
684 https://doi.org/10.1111/gbi.12243

685 Kacar B, Garcia AK, Anbar AD. 2020. Evolutionary History of Bioessential Elements Can Guide
686 the Search for Life in the Universe. *ChemBioChem*. doi: 10.1002/cbic.202000500.

687 Kaçar B, Gaucher EA. 2012. Towards the recapitulation of ancient history in the laboratory:
688 Combining synthetic biology with experimental evolution. *Artif Life*. 13:11-18. doi: 10.1162/978-
689 0-262-31050-5-ch002.

690 Kaiser JT, Hu Y, Wiig JA, Rees DC, Ribbe MW. 2011. Structure of Precursor-Bound NifEN: A
691 Nitrogenase FeMo Cofactor Maturase/Insertase. *Science*. 331:91-94. doi:
692 10.1126/science.1196954.

693 Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7:
694 improvements in performance and usability. *Mol Biol Evol*. 30:772-780. doi:
695 10.1093/molbev/mst010.

696 Kim CH, Newton WE, Dean DR. 1995. Role of the MoFe protein alpha-subunit histidine-195
697 residue in FeMo-cofactor binding and nitrogenase catalysis. *Biochemistry*. 34:2798-2808.

698 Kim J, Rees D. 1992. Structural models for the metal centers in the nitrogenase molybdenum-
699 iron protein. *Science*. 257:1677-1682. doi: 10.1126/science.1529354.

700 Lee H-I, Thrasher KS, Dean DR, Newton WE, Hoffman BM. 1998. ^{14}N Electron Spin−Echo
701 Envelope Modulation of the S = 3/2 Spin System of the *Azotobacter vinelandii* Nitrogenase
702 Iron−Molybdenum Cofactor. *Biochemistry*. 37:13370-13378. doi: 10.1021/bi980956a.

Li H. 2002. *Azotobacter vinelandii* Nitrogenase: Multiple Substrate-Reduction Sites and Effects of pH on Substrate Reduction and CO Inhibition. [Dissertation]. [Blacksburg, VA]: Virginia Polytechnic Institute and State University.

Luo G, Junium CK, Izon G, Ono S, Beukes NJ, Algeo TJ, Cui Y, Xie S, Summons RE. 2018. Nitrogen fixation sustained productivity in the wake of the Palaeoproterozoic Great Oxygenation Event. *Nat Commun*. 9:978. doi: 10.1038/s41467-018-03361-2.

McGlynn SE, Boyd ES, Peters JW, Orphan VJ. 2012. Classifying the metal dependence of uncharacterized nitrogenases. *Front Microbiol*. 3:419. doi: 10.3389/fmicb.2012.00419.

Mehta MP, Baross JA. 2006. Nitrogen fixation at 92 degrees C by a hydrothermal vent archaeon. *Science*. 314:1783-1786. doi: 10.1126/science.1134772.

Moser J, Brocker MJ. 2011. Enzymatic Systems with Homology to Nitrogenase. In: Ribbe M, editor. Nitrogen Fixation. Methods in Molecular Biology (Methods and Protocols): Humana Press. p. 67-77.

Mus F, Alleman AB, Pence N, Seefeldt LC, Peters JW. 2018. Exploring the alternatives of biological nitrogen fixation. *Metallomics*. 10:523-538. doi: 10.1039/c8mt00038g.

Mus F, Colman DR, Peters JW, Boyd ES. 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 CP, Mvondo DN. 2001. A possible nitrogen crisis for Archaean life due to reduced nitrogen fixation by lightning. *Nature*. 412:61-64. doi: Doi 10.1038/35083537.

Noda-Garcia L, Liebermeister W, Tawfik DS. 2018. Metabolite-Enzyme Coevolution: From Single Enzymes to Metabolic Pathways and Networks. *Annu Rev Biochem*. 87:187-216. doi: 10.1146/annurev-biochem-062917-012023.

North JA, Narrowe AB, Xiong W, Byerly KM, Zhao G, Young SJ, Murali S, Wildenthal JA, Cannon WR, Wrighton KC, et al. 2020. A nitrogenase-like enzyme system catalyzes methionine, ethylene, and methane biogenesis. *Science*. 369:1094-1098. doi: 10.1126/science.abb6310.

Ohno S. 1970. Evolution by gene duplication. Berlin: Springer-Verlag.

Parsons C, Stüeken EE, Rosen CJ, Mateos K, Anderson RE. 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.

Price MN, Dehal PS, Arkin AP. 2010. FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One*. 5:e9490. doi: 10.1371/journal.pone.0009490.

Raymond J, Siefert JL, Staples CR, Blankenship RE. 2004. The natural history of nitrogen fixation. *Mol Biol Evol*. 21:541-554. doi: 10.1093/molbev/msh047.

Seefeldt LC, Yang ZY, Lukoyanov DA, Harris DF, Dean DR, Raugei S, Hoffman BM. 2020. Reduction of Substrates by Nitrogenases. *Chem Rev*. 120:5082-5106. doi: 10.1021/acs.chemrev.9b00556.

740 Sickerman NS, Tanifuji K, Hu Y, Ribbe MW. 2017. Synthetic Analogues of Nitrogenase
741 Metallocofactors: Challenges and Developments. *Chemistry*. 23:12425-12432. doi:
742 10.1002/chem.201702496.

743 Silver WS, Postgate JR. 1973. Evolution of asymbiotic nitrogen fixation. *J Theor Biol*. 40:1-10.
744 doi: 10.1016/0022-5193(73)90160-4.

745 Solomon JB, Lee CC, Jasniewski AJ, Rasekh MF, Ribbe MW, Hu Y. 2020. Heterologous
746 Expression and Engineering of the Nitrogenase Cofactor Biosynthesis Scaffold NifEN. *Angew*
747 *Chem Int Ed Engl*. 59:6887-6893. doi: 10.1002/anie.201916598.

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

750 Staples CR, Lahiri S, Raymond J, Von Herbulis L, Mukhophadhyay B, Blankenship RE. 2007.
751 Expression and association of group IV nitrogenase NifD and NifH homologs in the non-
752 nitrogen-fixing archaeon *Methanocaldococcus jannaschii*. *J Bacteriol*. 189:7392-7398. doi:
753 10.1128/JB.00876-07.

754 Stueken EE, Buick R, Guy BM, Koehler MC. 2015. Isotopic evidence for biological nitrogen
755 fixation by molybdenum-nitrogenase from 3.2 Gyr. *Nature*. 520:666-669. doi:
756 10.1038/nature14180.

757 Stüeken EE, Kipp MA, Koehler MC, Buick R. 2016. The evolution of Earth's biogeochemical
758 nitrogen cycle. *Earth-Sci Rev*. 160:220-239. doi: 10.1016/j.earscirev.2016.07.007.

759 Tanifuji K, Lee CC, Ohki Y, Tatsumi K, Hu Y, Ribbe MW. 2015. Combining a Nitrogenase
760 Scaffold and a Synthetic Compound into an Artificial Enzyme. *Angew Chem Int Ed Engl*.
761 127:14228-14231. doi: 10.1002/ange.201507646.

762 Thornton JW. 2004. Resurrecting ancient genes: experimental analysis of extinct molecules.
763 *Nat Rev Genet*. 5:366-375. doi: 10.1038/nrg1324.

764 Vermeij GJ. 2006. Historical contingency and the purported uniqueness of evolutionary
765 innovations. *Proc Natl Acad Sci USA*. 103:1804-1809. doi: 10.1073/pnas.0508724103.

766 Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 24:1586-
767 1591. doi: 10.1093/molbev/msm088.

768 Zheng K, Ngo PD, Owens VL, Yang X-p, Mansoorabadi SO. 2016. The biosynthetic pathway of
769 coenzyme F430 in methanogenic and methanotrophic archaea. *Science*. 354:339-342. doi:
770 10.1126/science.aag2947.

771

772

773

774

775 **SUPPLEMENTARY INFORMATION**
776
777
778 **Table S1.** Jensen-Shannon (J-S) distances between nitrogenases and maturases at putative
779 functionally divergent sites (see main text for definition of divergent sites), calculated for sequence
780 alignments used to build Tree-1 to Tree-4. Site numbers from aligned *Azotobacter vinelandii*
781 nitrogenase NifD protein. Cells are colored grey if the corresponding site is not defined as
782 functionally divergent for the given alignment.
783

	J-S distance			
Site	Tree-1	Tree-2	Tree-3	Tree-4
15	0.7304	0.7358	0.73	
16	0.5438	0.5637	0.5442	
17	0.5687	0.5704	0.5706	0.566
19				0.6098
20	0.5462	0.5772	0.5461	0.6957
21	0.5456	0.5775	0.5455	0.6815
22	0.7661	0.767	0.7659	
23	0.6701	0.6889	0.6699	0.5829
24				0.54
25	0.5837	0.5981	0.5859	0.6326
26	0.6058	0.6209	0.605	0.615
27				0.5254
31	0.5658	0.5953	0.5652	
34	0.5509	0.5666	0.5499	
41	0.5423	0.582	0.5419	0.538
43				0.6596
44				0.5749
45				0.6333
46	0.6614	0.6726	0.6608	
47				0.6297
48	0.6955	0.7535	0.6953	0.6234
49	0.6484	0.6464	0.6481	0.6647
50				0.5262
51	0.57	0.5928	0.5701	0.535
52		0.5784		0.5347
56	0.6579	0.6882	0.6574	0.6462
57	0.7084	0.7332	0.708	0.687
58				0.5344
65	0.6453	0.6409	0.6448	0.6448
68	0.5934	0.6489	0.5929	0.5929
69	0.797	0.796	0.7969	0.7969
72	0.5676		0.5674	0.5674
73	0.7624	0.779	0.7623	0.7623
76	0.659	0.7087	0.6585	0.6585

78	0.5797	0.6037	0.5791	0.5791
89	0.5643	0.5959	0.564	0.5626
90	0.6805	0.7182	0.68	0.6782
94	0.711	0.717	0.7106	0.5681
95	0.681	0.6776	0.6805	0.5629
97	0.7381	0.7687	0.7378	0.5995
98	0.7749	0.7797	0.7747	0.7077
99		0.5579		0.5937
102	0.6329	0.6742	0.6324	
104	0.6927	0.7335	0.6924	
106	0.6831	0.7373	0.6827	
108	0.7818	0.782	0.7817	
109	0.7092	0.7157	0.7088	0.7096
110	0.7558	0.755	0.7556	0.7391
111	0.6264	0.6874	0.6258	0.6334
112	0.5977	0.602	0.5974	0.6044
113	0.6909	0.6922	0.6905	0.6776
118	0.5922	0.6003	0.5916	0.5916
139	0.5519	0.5696	0.5512	0.5513
141	0.6037	0.631	0.6033	0.6045
142	0.6541	0.6751	0.6538	0.6535
143	0.7407	0.7854	0.7404	
144	0.6035	0.631	0.6032	
149	0.6813	0.7186	0.6811	0.6811
155	0.7239	0.7618	0.7236	0.7236
167	0.6155	0.637	0.6199	0.6199
182	0.6057	0.629	0.6051	0.6051
187	0.745	0.7724	0.7447	
188	0.7239	0.7789	0.7236	
189	0.7818	0.7956	0.7817	0.7821
190	0.5563	0.5662	0.5552	0.5552
191	0.7429	0.8003	0.7425	0.7425
192	0.657	0.6556	0.6563	0.6563
195	0.7191	0.7697	0.7187	0.7187
196	0.7834	0.7817	0.7833	0.7833
197	0.5393	0.5559	0.5412	0.5412
199	0.7649	0.7801	0.7647	0.7647
200	0.5421	0.5689	0.5431	0.5431
205	0.57	0.576	0.5709	0.573
216	0.6118	0.606	0.6121	
231	0.5517		0.5506	0.5512
232	0.6688	0.6949	0.6725	0.6712
234	0.6597	0.6785	0.6591	0.6591
235	0.547	0.5947	0.546	0.546
237	0.545	0.5839	0.5443	0.5443

253				0.5439
274	0.6104	0.6484	0.6097	0.6097
276	0.6672	0.6705	0.6668	0.6668
277	0.7018	0.7167	0.7014	0.7014
279	0.5522		0.5524	0.5524
280	0.6949	0.7157	0.6945	0.6945
281	0.5782	0.5753	0.5774	0.5774
282	0.5504		0.5506	0.5506
285	0.553	0.5742	0.5547	0.5547
294	0.5653	0.6207	0.5645	0.5645
298	0.6738	0.7358	0.6735	0.6735
302		0.5632		
304	0.6519	0.6682	0.6519	0.6519
327	0.5514	0.5633	0.5507	0.5219
352	0.5876	0.6096	0.5909	0.5909
358	0.7199	0.7429	0.7196	0.7453
359	0.5804	0.6319	0.5801	0.5801
361	0.6183	0.6353	0.6189	0.6189
362	0.7864	0.7915	0.7863	0.7863
363	0.6387	0.6433	0.6385	0.6385
365	0.5913	0.6028	0.5905	0.562
379	0.5799	0.5916	0.58	0.58
380	0.6156	0.6136	0.6206	0.6206
381	0.7379	0.7954	0.7376	0.7376
383	0.7892	0.7879	0.7891	0.7891
385				0.5347
390	0.7294	0.7346	0.7291	0.6225
391	0.7297	0.7407	0.7293	
393	0.6666	0.6826	0.666	0.6329
403				0.6413
404	0.5889	0.6045	0.5882	
405				0.7308
406	0.6473	0.6811	0.6483	0.7155
407	0.571	0.5914	0.5713	0.5764
415	0.5565	0.5782	0.5558	0.5547
416	0.635	0.6562	0.635	0.6327
418	0.6885	0.7324	0.688	0.688
421	0.559	0.587	0.5584	0.5584
422	0.677	0.7234	0.6766	0.6766
423	0.5441		0.5445	0.5445
425	0.6596	0.6761	0.6592	0.6592
428	0.7035	0.7436	0.7032	0.684
431				0.5444
432	0.7118	0.7586	0.7116	0.7108
439	0.6033	0.6253	0.6035	0.6045

440	0.7852	0.7873	0.7851	0.7759
441	0.6988	0.7037	0.7004	0.7521
442	0.7998	0.8001	0.7998	
443	0.7556	0.7598	0.7554	0.6971
444	0.7883	0.7874	0.7883	0.7927
445	0.7894	0.7934	0.7893	0.6103
446	0.5442	0.5761	0.5441	0.7547
447	0.6367	0.7113	0.6365	0.6679
448	0.7598	0.7767	0.7597	0.7597
456		0.6125		0.5321
462	0.5893	0.6013	0.589	0.589
477		0.5634		

Table S2. Ancestral D-scores for putative functionally divergent sites between nitrogenases and maturases, calculated for sequence alignments used to build Tree-1 to Tree-4. Site numbers from aligned *Azotobacter vinelandii* nitrogenase NifD protein. See main text for definitions of ancestral nodes ("LCA": Last common ancestor). Positive values are colored red and negative values are colored blue. Cells are colored grey if the corresponding site is not defined as functionally divergent for the given alignment.

	Nitrogenase/maturase LCA 1				Nitrogenase/maturase LCA 2				Nitrogenase LCA			
Site	Tree-1	Tree-2	Tree-3	Tree-4	Tree-1	Tree-2	Tree-3	Tree-4	Tree-1	Tree-2	Tree-3	Tree-4
15	-0.48	-0.16	-0.56		-0.62	-0.15	-0.62		-0.59	0.09	-0.57	
16	0.11	0.35	-0.12		0.07	0.38	-0.15		0.11	0.38	0.24	
17	-0.30	-0.14	-0.09	-0.41	-0.26	0.11	-0.08	-0.22	0.29	0.45	0.28	0.10
19				-0.43				-0.47				0.16
20	-0.59	-0.61	-0.59	-0.44	-0.59	-0.61	-0.63	-0.42	-0.59	-0.60	-0.59	0.35
21	-0.39	-0.61	0.38	-0.44	-0.57	-0.60	-0.63	-0.40	-0.60	-0.50	-0.59	0.07
22	-0.67	-0.70	-0.67		-0.67	-0.67	-0.72		-0.69	0.07	-0.68	
23	-0.32	-0.32	-0.33	-0.18	-0.31	-0.31	-0.33	-0.11	-0.32	0.44	-0.32	0.38
24				-0.28				-0.30				0.18
25	-0.38	-0.37	-0.03	-0.24	-0.44	-0.35	-0.26	-0.28	-0.28	-0.20	-0.26	0.01
26	-0.31	-0.10	-0.24	-0.24	-0.08	0.16	-0.04	-0.27	0.16	0.25	0.15	0.21
27				-0.25				-0.27				0.07
31	0.01	0.29	-0.01		0.27	0.45	0.24		0.45	0.45	0.45	
34	0.02	0.21	-0.01		0.17	0.45	0.20		0.41	0.43	0.40	
41	-0.39	0.30	-0.32	-0.61	0.24	0.21	0.23	-0.61	0.22	0.21	0.22	-0.65
43				-0.48				-0.48				0.09
44				-0.44				-0.44				0.18
45				-0.43				-0.43				0.15
46	-0.47	0.05	-0.44		0.03	0.11	0.24		0.24	0.13	0.18	
47				-0.21				0.38				0.58
48	-0.34	-0.44	-0.03	-0.09	-0.41	-0.62	-0.36	-0.39	-0.57	-0.56	-0.57	-0.36
49	-0.02	0.12	-0.03	-0.08	0.03	0.37	-0.05	0.12	0.44	0.43	0.41	0.45
50				-0.12				-0.15				0.22
51	0.16	0.11	-0.27	-0.08	0.13	0.20	-0.40	0.02	0.20	0.24	-0.21	0.15
52		0.42		-0.07		0.45		0.39		0.56		0.36

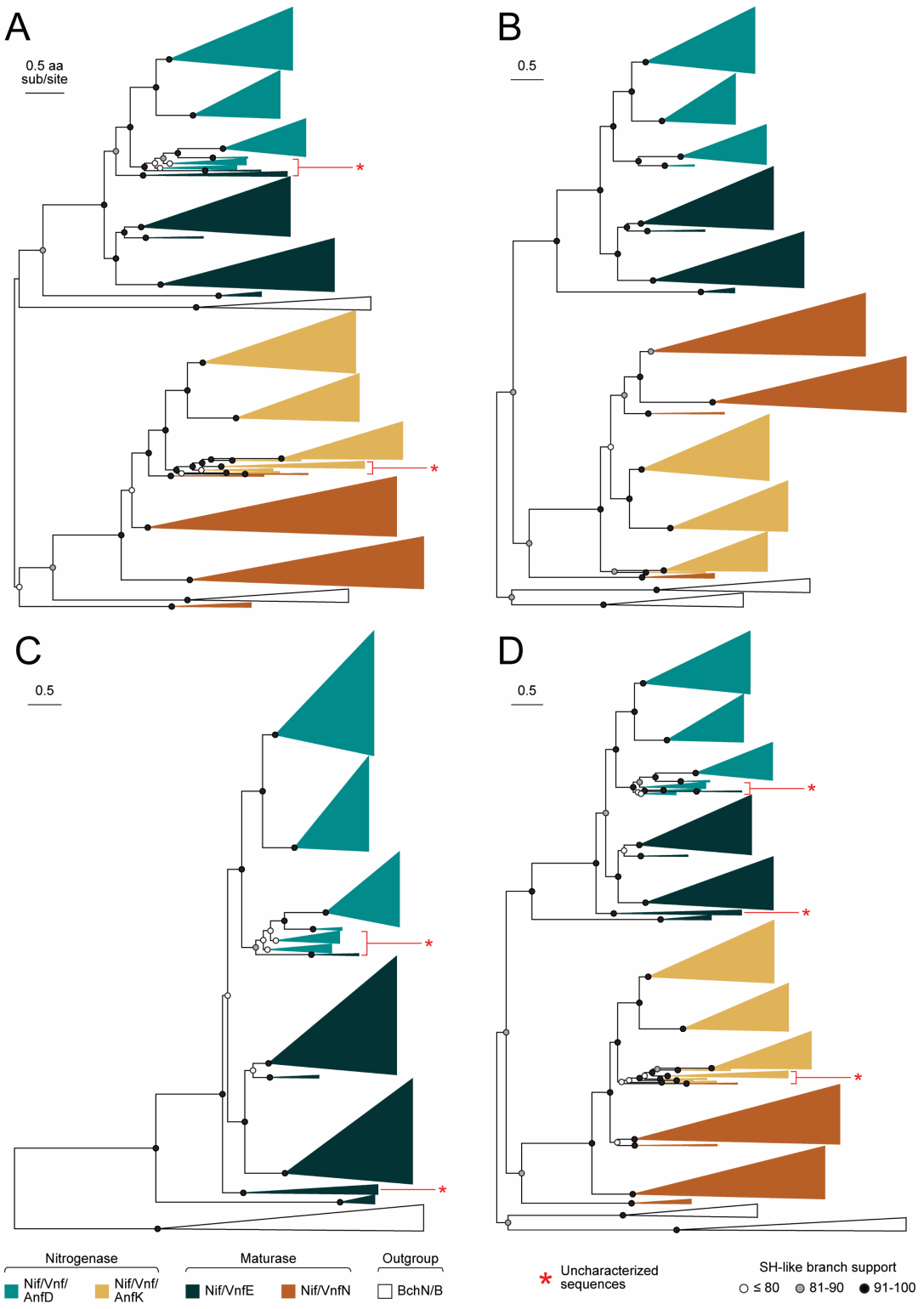
56	0.30	0.38	0.34	-0.19	0.28	0.23	0.29	-0.01	0.23	0.30	0.25	-0.01
57	-0.02	0.03	-0.03	0.44	-0.04	0.06	0.09	0.24	0.20	0.31	0.21	0.22
58				0.39				0.36				0.34
65	0.40	0.42	0.41	0.16	0.39	0.42	0.28	0.22	0.39	0.45	0.32	0.44
68	-0.02	-0.05	0.00	-0.05	-0.06	0.16	-0.28	-0.28	-0.28	0.68	-0.28	-0.28
69	-0.69	-0.71	-0.69	-0.69	-0.69	-0.71	-0.69	-0.69	-0.69	0.43	-0.69	0.44
72	-0.19	0.03	0.00	-0.49	-0.49	0.10	-0.62	-0.59	-0.59	0.50	-0.59	-0.60
73	-0.02	0.43	0.00	-0.03	-0.02	0.38	0.19	0.00	0.02	0.37	0.35	0.51
76	0.40	-0.38	0.41	0.45	0.40	-0.45	0.55	0.36	0.36	0.10	0.36	0.35
78	-0.44	-0.33	-0.02	-0.37	-0.45	-0.34	-0.36	-0.48	-0.48	-0.22	-0.40	-0.10
89	-0.37	0.00	-0.36	-0.37	-0.36	0.00	-0.40	-0.35	-0.35	0.14	-0.35	-0.36
90	0.21	0.11	0.05	0.20	0.20	0.06	0.24	0.19	0.20	0.08	0.20	0.19
94	-0.01	-0.17	0.09	0.03	-0.01	-0.07	0.14	-0.21	-0.02	0.38	0.02	-0.28
95	-0.21	0.70	0.04	0.29	-0.23	0.72	-0.12	0.29	-0.15	0.71	-0.14	0.29
97	-0.03	-0.06	0.00	0.17	0.00	0.06	0.24	0.28	0.25	0.37	0.13	0.42
98	-0.24	0.04	-0.01	0.25	-0.29	-0.03	-0.52	0.15	0.01	-0.01	-0.18	0.28
99				-0.33				-0.29				-0.23
102	0.15	-0.21	0.12		-0.19	-0.56	-0.11		-0.43	-0.40	-0.51	
104	-0.61	-0.09	-0.57		-0.66	-0.07	-0.69		-0.57	0.36	-0.57	
106	0.34	-0.51	0.35		-0.54	-0.32	-0.52		-0.17	-0.31	-0.42	
108	0.00	0.04	0.08		-0.73	0.15	-0.76		-0.74	0.39	-0.71	
109	-0.23	-0.34	-0.01	-0.04	-0.29	-0.45	-0.13	-0.05	-0.20	-0.12	-0.17	0.07
110	-0.22	-0.37	0.01	-0.06	-0.50	-0.32	-0.46	-0.06	-0.30	0.20	-0.30	-0.06
111	-0.19	-0.23	0.00	-0.12	-0.53	-0.22	-0.17	-0.16	-0.49	-0.07	-0.43	-0.03
112	-0.10	-0.12	-0.02	-0.07	-0.10	0.01	-0.14	-0.09	0.20	0.34	0.09	0.36
113	0.04	0.04	0.00	0.02	0.06	0.09	0.22	0.06	0.17	0.10	0.14	0.18
118	-0.42	-0.30	-0.47	-0.44	-0.44	-0.29	-0.54	-0.42	-0.44	-0.09	-0.42	0.24
139	-0.27	-0.02	-0.05	-0.04	-0.25	0.00	-0.13	-0.17	-0.19	0.23	-0.21	0.16
141	-0.22	-0.26	-0.14	-0.32	-0.24	-0.25	-0.21	-0.20	-0.09	-0.32	0.06	-0.21
142	-0.26	-0.32	0.00	-0.27	-0.27	-0.26	0.07	0.26	0.45	0.42	0.33	0.50
143	-0.63	-0.71	-0.62		-0.63	-0.67	-0.75		0.30	0.31	-0.37	
144	-0.42	-0.39	-0.09		-0.44	-0.19	-0.34		-0.21	-0.11	-0.11	

149	-0.60	-0.63	-0.59	-0.62	-0.59	-0.61	-0.61	-0.58	-0.58	-0.52	-0.58	-0.59
155	0.16	0.08	0.02	0.08	0.15	0.05	0.06	0.08	0.24	0.10	0.05	0.14
167	-0.64	-0.60	-0.13	-0.61	-0.62	-0.54	-0.72	-0.48	-0.49	-0.16	-0.50	-0.10
182	0.23	0.07	0.18	0.11	0.15	0.07	-0.01	0.07	0.07	0.07	0.07	0.07
187	-0.57	-0.65	-0.57		-0.58	-0.66	-0.40		0.37	0.35	0.27	
188	-0.22	-0.13	-0.01		-0.19	-0.01	0.33		0.71	0.72	0.71	
189	-0.61	-0.68	-0.02	-0.65	-0.61	-0.69	-0.02	0.07	0.32	0.25	0.33	0.48
190	0.38	0.38	-0.15	0.44	0.40	0.41	0.24	0.65	0.65	0.65	0.65	0.65
191	-0.02	-0.02	0.00	-0.02	-0.02	-0.07	-0.03	0.00	0.00	0.00	-0.01	0.00
192	0.09	0.07	-0.02	0.08	0.19	0.57	0.37	0.61	0.72	0.71	0.71	0.71
195	-0.13	-0.07	0.00	-0.14	-0.09	0.68	0.60	0.73	0.73	0.73	0.72	0.72
196	-0.06	-0.03	-0.04	-0.06	-0.07	0.29	-0.10	-0.14	-0.11	0.75	-0.27	0.75
197	-0.27	-0.13	-0.18	-0.05	-0.17	-0.02	0.06	0.25	0.24	0.31	0.25	0.25
199	-0.04	-0.09	0.00	-0.05	-0.07	-0.06	-0.06	-0.07	-0.04	0.75	0.01	0.73
200	-0.15	-0.15	-0.23	-0.13	-0.13	0.17	-0.10	0.00	0.03	0.24	0.04	0.16
205	0.14	0.14	0.12	-0.06	0.14	0.04	0.07	0.01	0.01	0.07	0.02	0.02
216	-0.43	-0.30	-0.42		-0.47	-0.31	-0.53		-0.46	0.06	-0.44	
231	-0.26		-0.11	-0.28	-0.16		-0.11	-0.09	-0.01		-0.10	0.49
232	-0.14	0.05	-0.18	-0.03	-0.20	0.01	-0.51	-0.48	-0.42	0.08	-0.50	-0.11
234	0.70	0.70	0.00	0.70	0.70	0.70	0.71	0.70	0.70	0.70	0.70	0.70
235	-0.21	-0.25	0.00	-0.22	-0.21	-0.16	-0.28	-0.12	-0.12	-0.17	-0.12	-0.12
237	-0.09	-0.09	0.10	-0.08	-0.08	-0.06	0.03	0.14	0.14	-0.05	0.16	0.13
253				-0.19				0.17				0.17
274	-0.02	0.02	0.00	-0.02	-0.02	0.22	0.19	0.18	0.17	0.28	0.17	0.17
276	0.03	0.06	0.00	0.02	0.01	0.04	0.00	0.01	0.03	0.08	0.01	0.08
277	-0.10	-0.09	0.00	-0.07	-0.12	0.06	-0.18	0.23	0.11	0.60	-0.12	0.66
279	0.10		-0.13	-0.05	0.09		-0.18	-0.03	0.11		-0.09	0.09
280	0.00	0.02	0.01	-0.07	-0.03	0.14	-0.17	0.01	0.02	0.18	-0.06	0.02
281	-0.26	-0.31	-0.17	-0.33	-0.19	0.65	-0.21	0.66	0.69	0.68	0.69	0.68
282	-0.17		0.01	-0.20	-0.27		-0.30	-0.33	-0.23		-0.31	0.33
285	-0.29	-0.36	-0.05	-0.10	-0.38	-0.36	-0.47	-0.12	-0.09	-0.11	-0.34	0.15
294	-0.41	-0.50	-0.01	-0.40	-0.40	-0.25	-0.41	-0.18	-0.17	-0.21	-0.17	-0.18

298	-0.01	-0.12	-0.01	-0.12	-0.11	-0.10	-0.16	-0.55	-0.58	0.50	-0.56	-0.58
302		-0.46				-0.48				-0.41		
304	-0.05	-0.19	0.03	0.02	-0.07	-0.38	0.10	-0.09	-0.09	-0.31	-0.10	-0.09
327	0.00	-0.12	0.00	-0.27	-0.03	0.06	0.44	0.43	0.63	0.64	0.50	0.54
352	-0.22	-0.07	0.26	-0.24	-0.23	0.00	-0.17	-0.20	-0.20	0.40	-0.16	0.28
358	-0.15	-0.18	-0.03	0.14	-0.15	-0.17	0.17	0.27	0.21	-0.02	0.23	0.28
359	0.01	-0.10	0.00	0.40	0.04	-0.27	0.59	0.60	0.65	-0.07	0.52	0.50
361	-0.24	-0.32	-0.13	-0.28	-0.28	-0.09	-0.29	-0.13	-0.05	-0.09	-0.05	-0.05
362	-0.32	-0.40	0.00	-0.39	-0.48	-0.46	-0.63	-0.64	-0.63	0.75	-0.59	0.75
363	-0.37	-0.39	-0.28	-0.40	-0.38	-0.34	-0.38	-0.20	-0.24	-0.33	-0.15	0.14
365	-0.15	-0.18	-0.09	-0.18	-0.18	-0.23	-0.21	-0.15	-0.18	-0.21	-0.16	-0.13
379	-0.17	-0.24	-0.17	-0.21	-0.18	-0.25	-0.28	-0.23	-0.19	-0.28	-0.18	-0.18
380	0.17	0.21	0.00	0.04	0.23	0.39	0.24	0.35	0.34	0.35	0.35	0.23
381	-0.03	0.02	0.00	-0.01	-0.03	0.09	0.47	0.74	0.74	0.74	0.73	0.73
383	-0.27	0.47	0.25	-0.17	-0.28	0.53	-0.03	-0.15	-0.12	0.74	0.08	0.77
385				0.00				0.04				0.30
390	0.09	0.22	0.02	-0.03	0.18	0.28	0.11	0.01	-0.68	0.24	-0.68	-0.02
391	-0.64	-0.02	-0.62		-0.67	-0.02	-0.73		-0.64	0.37	-0.64	
393	-0.36	0.00	-0.32	-0.30	-0.24	0.17	0.10	-0.22	0.17	0.27	0.18	0.14
403				0.60				0.55				0.04
404	-0.19	-0.19	-0.02		-0.19	-0.11	-0.31		-0.11	0.00	-0.11	
405				-0.65				-0.66				0.29
406	0.01	-0.01	0.02	-0.64	-0.03	-0.01	-0.03	-0.65	-0.03	0.02	-0.03	0.08
407	0.38	0.37	0.00	0.45	0.40	0.35	0.01	0.03	0.21	0.38	0.04	0.08
415	-0.11	-0.12	-0.01	-0.12	-0.10	-0.05	0.13	0.05	0.06	-0.04	0.03	0.05
416	-0.02	-0.11	0.41	0.09	-0.04	0.08	0.11	0.03	0.03	0.46	0.03	0.04
418	0.68	0.71	0.66	0.69	0.69	0.67	0.69	0.66	0.66	0.66	0.66	0.66
421	-0.10	0.04	0.26	0.01	-0.27	-0.01	0.03	0.03	0.04	0.15	0.04	0.05
422	-0.52	-0.40	-0.42	-0.56	-0.57	-0.51	-0.66	-0.40	-0.44	-0.31	-0.35	-0.47
423	-0.30		0.11	-0.29	-0.30		-0.38	-0.32	-0.32		-0.05	0.18
425	-0.08	-0.02	0.03	-0.04	-0.08	0.10	0.08	-0.04	-0.04	0.36	0.09	0.29
428	0.40	0.39	0.00	0.40	0.40	0.40	0.21	0.39	0.39	0.40	0.39	0.39

431				-0.52				-0.43				-0.49
432	0.17	0.07	0.00	0.10	-0.04	0.01	-0.03	0.00	0.00	0.08	0.00	-0.01
439	-0.14	-0.31	-0.01	-0.12	-0.15	-0.43	-0.18	-0.11	-0.11	-0.31	-0.15	-0.11
440	-0.46	-0.57	-0.01	0.03	-0.48	-0.56	-0.46	0.03	-0.04	-0.05	-0.39	0.08
441	0.03	0.12	0.00	0.02	0.11	0.02	0.01	-0.01	-0.01	0.03	-0.05	0.15
442	-0.61	-0.46	0.00		-0.56	-0.39	-0.40		0.27	0.73	-0.29	
443	0.22	0.58	0.00	0.18	0.23	0.41	0.16	0.20	0.15	0.39	0.07	0.38
444	0.02	0.36	0.00	0.25	0.01	0.38	-0.08	0.07	-0.01	0.38	-0.02	0.39
445	-0.34	-0.19	-0.03	0.04	-0.35	-0.18	-0.16	0.13	-0.08	0.28	-0.15	0.15
446	-0.58	-0.63	-0.57	-0.29	-0.58	-0.63	-0.57	-0.33	-0.59	-0.63	-0.57	-0.40
447	0.08	0.12	0.00	0.10	0.12	0.14	0.11	0.12	0.16	0.15	0.05	0.07
448	-0.14	-0.20	-0.02	-0.09	-0.14	-0.08	-0.08	0.00	0.04	0.75	0.00	0.70
456		0.00		0.05		-0.01		0.16		0.16		0.15
462	-0.10	-0.07	-0.01	-0.08	-0.12	0.15	-0.01	0.43	0.39	0.44	0.43	0.48
477		-0.37				-0.27				-0.20		

3



4
5
6 **Fig. S1.** Alternate maximum likelihood phylogenies built from nitrogenase and maturase protein
7 sequences (see main text for tree reconstruction methods): (*A*) Tree-1, (*B*) Tree-2, (*C*) Tree-3,
8 and (*D*) Tree-4. Outgroup clade widths are not to scale.

9

10