DEFINING THE ECOLOGICAL INTERACTIONS THAT DROVE THE EVOLUTION OF BIOLOGICAL NITROGEN FIXATION

by

Trinity Lynn Hamilton

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

Doctor of Philosophy

in

Biochemistry

MONTANA STATE UNIVERSITY
Bozeman, Montana

April 2012

©COPYRIGHT

by

Trinity Lynn Hamilton

2012

All Rights Reserved

APPROVAL

of a dissertation submitted by

Trinity Lynn Hamilton

This dissertation has been read by each member of the dissertation committee and has been found to be satisfactory regarding content, English usage, format, citation, bibliographic style, and consistency and is ready for submission to The Graduate School.

Dr. John W. Peters

Approved for the Department of Chemistry and Biochemistry

Dr. Bern Kohler

Approved for The Graduate School

Dr. Carl A. Fox

STATEMENT OF PERMISSION TO USE

In presenting this dissertation in partial fulfillment of the requirements for a doctoral degree at Montana State University, I agree that the Library shall make it available to borrowers under rules of the Library. I further agree that copying of this dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for extensive copying or reproduction of this dissertation should be referred to ProQuest Information and Learning, 300 North Zeeb Road, Ann Arbor, Michigan 48106, to whom I have granted "the exclusive right to reproduce and distribute my dissertation in and from microform along with the non-exclusive right to reproduce and distribute my abstract in any format in whole or in part."

Trinity L. Hamilton

April 2012

ACKNOWLEDGEMENTS

I am very grateful to my advisor, Professor John Peters, for giving me the opportunity to study and conduct research on a number of exciting projects and for providing guidance throughout my graduate studies. I would also like to thank Dr. Eric Boyd for mentoring, patience, and encouragement as well as numerous ideas and insight. I would like to thank the members of my committee, Dr. Joan Broderick, Dr. Martin Teintze, Dr. Brian Bothner, Dr. Mike Franklin, Dr. Tim McDermott and Dr. Yves Idzerda for useful discussion and direction.

I would like to extend many thanks to the members of the Peters' lab for all of their contributions to various aspects of research problems and discussions as well as maintaining an enjoyable and supportive research environment. In addition, I wish to acknowledge Dr. Ross Taylor for teaching me many valuable skills and encouraging research creativity.

I would also like to thank many friends around the Gallatin Valley and elsewhere for their continued support. I would especially like to thank my family, my parents Dan and Pam, my sister Trista and her fiancé Ryan, my brother Travis and his wife Laura and Jeff for always believing in me and encouraging me. Special thanks to Rylee, Rance, and Tate, the ones who remind me what life is all about, having fun (and toys).

TABLE OF CONTENTS

1. INTRODUCTION 1

Biological Nitrogen Fixation 3
The Nitrogenase Enzyme 6
The Fe Protein 8
The MoFe Protein 11
Biosynthesis of Nitrogenase 13
Distribution of Biological Nitrogen Fixation 17
Yellowstone National Park—Abundant Geochemical Gradients 18
Microbial Diversity in Extreme Environments of Yellowstone National Park 20
Research Directions 23
Literature Cited 28

2. EXAMINING THE ORIGIN AND EVOLUTION OF BIOLOGICAL NITROGEN FIXATION 42

Contribution of Authors and Co-Authors 42
Manuscript Information Page 43
Introduction 44
Materials and Methods 49
Phylogenetic Analysis 49
Structural Analyses 50
Statistical Analyses 51
Vnf/Anf/NifDKEN and NifB Phylogenetic Analyses 51
Composite Vnf/Anf/NifDKEN and NifB Phylogenetic Analyses 53
Evolutionary Rate Analysis 53
Results and Discussion 55
An Alternative Path for the Evolution of Biological Nitrogen Fixation 55
A Late Methanogenic Origin for Molybdenum-dependent Nitrogenase 62
Origin of Nif 62
Temporal Relationships Between the Emergence of Nif and Oxygenic Photosynthesis 69
Conclusions 73
Acknowledgements 75
Literature Cited 76

3. ENVIRONMENTAL CONSTRAINTS UNDERPIN THE DISTRIBUTION AND PHYLOGENETIC DIVERSITY OF *NIFH* IN THE YELLOWSTONE GEOTHERMAL COMPLEX 83

TABLE OF CONTENTS CONTINUED

Contribution of Authors and Co-Authors....83
Manuscript Information Page....84
Abstract....85
Introduction....86
Methods....89
Sample Site Description....89
Sample Collection and DNA Extraction....89
nifH Primer Design and PCR Amplification....90
Chemical Analyses....92
Phylogenetic Analysis....93
Community Ecological Analyses....94
Nucleotide Sequence Accession Numbers....96
Results....96
Distribution of *nifH* in the YNP Geothermal Complex....96
Phylogenetic Diversity of *nifH* Assemblages....99
Discussion....105
Acknowledgements....108
Literature Cited....110

4. BIOLOGICAL NITROGEN FIXATION IN ACIDIC HIGH-TEMPERATURE GEOTHERMAL SPRINGS IN YELLOWSTONE NATIONAL PARK, WYOMING....115

Contribution of Authors and Co-Authors....115
Manuscript Information Page....116
Abstract....117
Introduction....118
Materials and Methods....119
Sample Collection and DNA Extraction....119
nifH Amplification and Clone Library Construction....120
Chemical Analyses....121
Nitrogenase Activity Assays....122
$^{15}N_2$ Uptake Analyses....124
Partial Diazotroph Enrichment Culture....125
Quantitative PCR (qPCR)....126
Phylogenetic Analysis....128
Results and Discussion....129
nifH Distribution and Diversity....130
Potential Biological N_2 Reduction Rates....133
Isotopic Studies of Biological N_2 Fixation....137
Enrichment of Thermoacidiphilic Diazotrophic sp. NG4 HFS....140

TABLE OF CONTENTS CONTINUED

Ecological Relevance of NG4 HFS to Acidic High Temperature YNP Spring Communities 142
Conclusions 143
Acknowledgements 145
Literature Cited....146

5. TRANSCRIPTIONAL PROFILING OF NITROGEN FIXATION IN *AZOTOBACTER VINELANDII* 150

Contribution of Authors and Co-Authors....150
Manuscript Information Page 152
Abstract 153
Introduction 154
Materials and Methods 157
Strains and Growth Conditions....157
Preparation of Total RNA, cDNA Library Constructions and SOLiD™ 157
Data Processing and Analysis....158
Quantitative RT-PCR....158
Results and Discussion....159
Expression of *nif* Genes Under Mo-dependent Nitrogen Fixing Conditions When Compared to Non-diazotrophic Growth 161
Differences in Global Patterns of Gene Expression Implicated During Mo-dependent Diazotrophic Growth When Compared to Non-diazotrophic Growth 165
Transcription Profiles in Cells Expressing Alternative Nitrogenases....169
Electron Transport for Mo-, V-, and Fe-only Nitrogenase Dependent Diazotrophic Growth 174
Regulation of Gene Expression During Diazotrophic Growth 175
Differences in Global Patterns of Gene Expression Implicated During Mo-independent Nitrogen Fixing Growth 177
Evolutionary Implications....179
Conclusion....180
Acknowledgements 181
Literature Cited....183

6. DIFFERENTIAL ACCUMULATION OF *NIF* STRUCTURAL GENE MRNA IN *AZOTOBACTER VINELANDII* 189

Contribution of Authors and Co-Authors....189
Manuscript Information Page 190

TABLE OF CONTENTS CONTINUED

Abstract 191
Acknowledgements 199
Literature Cited.....200

7. CONCLUDING REMARKS.....202

APPENDICES 207

APPENDIX A: Supporting Information for Chapter 2.....208
APPENDIX B: Supporting Information for Chapter 3.....244
APPENDIX C: Supporting Information for Chapter 4.....253
APPENDIX D: Supporting Information for Chapter 5.....257

LITERATURE CITED 294

LIST OF TABLES

Table Page

3.1. Physical and chemical data for YNP thermal features where *nifH* clone libraries were constructed and sequenced 97

3.2. Clone library statistics for the 13 geothermal spring environments analyzed 100

3.3 Phylogenetic diversity metrics for the thirteen *nifH* assemblages 101

3.4. Model ranking using ΔAICc and Mantel correlation coefficients (R^2) with Rao phylogenetic distance of *nifH* serving as the response variable 102

4.1. Biological, physical, and chemical data for YNP thermal features examined in the present study 129

4.2. Phylogenetic diversity metrics for the sixe *nifH* assemblages 130

4.3. Acetylene reduction rates 135

4.4. $^{15}N_2$ incorporation assays 139

4.5. Abundance of phylotype NG4 HFS *nifH* in YNP geothermal spring environments as assessed using qPCR 142

LIST OF FIGURES

Figure Page

1.1. Ribbons representation of NifF (*nif*-specific flavodoxin), the Fe protein, and the MoFe protein from *A. vinelandii* with the flow of electrons shown from flavodoxin to the Fe protein and from the Fe protein to the MoFe protein 5

1.2. The three state cycle of a single nitrogenase Fe protein – MoFe protein complex association and dissociation and electron transfer.....7

1.3. Ribbon representation of the overall fold of the *A. vinelandii* Fe protein dimer.....9

1.4. Ribbon representation of the two nucleotide dependent Switch regions and the phosphate binding 'P loop' region of the Fe protein 10

1.5. Ribbons representation of the nitrogen MoFe protein from *A. vinelandii* viewed down the twofold axis of symmetry that relates each equivalent dimer of the $\alpha_2\beta_2$ heterotetramer 11

1.6. FeMo-co and P cluster binding sites in the MoFe protein 13

1.7. Model for the biosynthesis of FeMo-co 14

1.8. Ribbons representation of NifEN from *A. vinelandii*.....16

1.9. Organization of nitrogen fixation (*nif*) genes in *A. vinelandii* and *M. maripaludis* 17

1.10. Map of Yellowstone National Park 21

2.1 Bayesian inferred phylogenetic tree of concatenated HDK homologs 56

2.2 Bayesian inferred phylogenetic reconstruction of Anf/Vnf/NifD, BchN, and NflD proteins 60

2.3 Model depicting the divergence of nitrogenase (NifD) and protochlorophyllide reductase (ChlN/BchN) from a NflD ancestor 61

2.4 Nitrogenase regulon from *Klebsiella pneumoniae* with arrows indicating the duplication events. Universal tree indicating lineages

LIST OF FIGURES CONTINUED

Figure Page

where full complements of nitrogenase proteins (HDKENB) were found. Bayesian consensus phylogenetic tree of Nif/Vnf/AnfD, K, E, and NPenalized-likelihood rate-smoothed chronogram of representative sequences for Vnf/Anf/NifDKEN and BchNBY 63

2.5 Bayesian consensus phylogenetic tree of Nif/Vnf/AnfD, K, E, and N proteins 66

2.6 Phylogenetic tree of the radical SAM domain of NifB and the genetic structure of the *nif* regulons for taxa presented in phylogenic tree 68

2.7 Penalized-likelihood rate-smoothed chronogram of representative sequences for Vnf/Anf/NifDKEN and BchNBYZ rooted with representatives from the VnfEN lineage. 71

3.1 Plot of the presence/absence of *nifH* in 64 mat and sediment samples collected from 4 geographic locations in the YNP geothermal complex plotted as a function of spring water temperature and pH 98

3.2 PCO analysis of *nifH* sequences from the 13 environments 104

4.1 Bayesian inferred phylogenetic reconstruction of *nifH* gene fragments recovered from YNP geothermal spring environments and from YNP geothermal spring enrichments 132

4.2 Acetylene reduction activity associated with sediment biomass sampled from NG4 136

4.3 Biomass atom %15N in assays 138

5.1 The global response of the *A. vinelandii* trancriptome during diazotrophic growth 160

5.2 Differential expression of proteins necessary for Mo-dependent diazotrophy 162

5.3 Differential expression of proteins necessary for Mo-independent nitrogen fixation 172

LIST OF FIGURES CONTINUED

Figure Page

6.1 Northern blot hybridization analysis of total RNA isolated from wild-type and mutant strains of *A. vinelandii*.....195

6.2 Identification of RNA secondary structures between the *nif* structural genes and northern blot hybridization analysis of total RNA isolated from mutant strain DJ81.....198

GLOSSARY

Nif – Mo-dependent nitrogenase
Vnf – V-dependent nitrogenase
Anf – Fe-only nitrogenase
LUCA – Last Universal Common Ancestor
LGT – Lateral gene transfer
YNP – Yellowstone National Park
PCR – Polymerase Chain Reaction
qPCR – quantitative Polymerase Chain Reaction
RT-PCR – Reverse Transcriptase Polymerase Chain Reaction
qRT-PCR – quantitative Reverse Transcriptase Polymerase Chain Reaction
SAM – S-adenosyl methionine
DPOR – Dark-operative protochlorophyllide reductase

ABSTRACT

All life requires fixed forms of nitrogen (N). On early Earth, fixed N was supplied through abiotic mechanisms, which became limiting to an expanding biome, precipitating the emergence of biological nitrogen fixation. Today, most biological nitrogen fixation is catalyzed by molybdenum (Mo)-dependent nitrogenase (Nif). Alternative forms of the enzyme contain either vanadium (V) or only iron (Fe) instead of Mo, but are only found in taxa that encode Nif. Geochemical evidence suggests Mo bioavailability was limited on the early Earth, leading to the hypothesis that alternative forms of nitrogenase are ancestral. Evidence presented here suggests that in fact Nif emerged first in a methanogenic archaeon. Previous studies revealed a widespread distribution of *nif* along geochemical gradients but little is known about the environmental conditions that drove its evolution. An analytical approach allowed examination of the role environment played in shaping the evolution of Nif across geochemical gradients in Yellowstone National Park. The distribution of *nifH* was widespread and not constrained by temperature or pH alone, but exhibited evidence of niche conservatism imposed by salinity, and seemed dispersal limited. Activity measurements in sediments collected from high-temperature acidic springs confirmed the potential for N_2 fixation in these environments. These data expand our understanding of the habitat range and environmental drivers of N_2 fixing organisms. In organisms that encode alternative nitrogenases, Nif is preferred for nitrogen fixation. In addition, the alternative forms of the enzyme do not encode the full suite enzymes necessary for assembling the active site metal cofactors. Presumably, the selective pressure driving the evolution of alternative nitrogenase would have been provided by Mo limitation. Transcriptome studies of a model organism which encodes all three forms of nitrogenase reveals the genes associated with expression of each nitrogenase and the interplay between systems that enables nitrogen fixation in the absence of Mo and fixed N. These analyses suggest the alternative nitrogenases would not function in the absence of Nif biosynthetic machinery and expression of nitrogenase is regulated by fixed N limitation and metal availability. The results presented here help elucidate the environmental conditions that have driven nitrogenase evolution, resulting in the extant enzyme.

CHAPTER 1

INTRODUCTION

Biologically available nitrogen is essential for all life. On early Earth, fixed forms of nitrogen such as ammonia (NH_3) and nitrous oxides (NO_x) were likely generated by abiotic processes such as lightning discharge breaking bonds of atmospheric dinitrogen (N_2). Biological nitrogen fixation, the conversion of dinitrogen gas to ammonia, is thought to have been integral to the evolution of life on Earth as abiotic sources of fixed nitrogen became limiting for the expanding biome. Atmospheric dinitrogen is the largest sink of nitrogen on Earth today; however most organisms are unable to utilize it in this inert form. The Haber-Bosch process provides the industrial means to convert dinitrogen to a biologically available nitrogen species, NH_3. This reaction is inefficient and energy costly, taking place under 250 atm and at temperatures of 450-500°C, with ammonia yields of 10-20%. Nitrogen fixation by prokaryotes produces the same result, but at or near physiological temperatures and pressure and accounts for nearly two thirds of the fixed nitrogen on Earth today (*1*). Biological nitrogen fixation plays a pivotal role in the global nitrogen cycle by reducing atmospheric nitrogen to a biologically available form according to the following reaction:

$$N_2 + 8H^+ + 8e^- + 16MgATP \rightarrow 2NH_3 + H_2 + 16MgADP + 16Pi$$

Simulations of the Archaean atmosphere indicate that a lack of abiotic sources of fixed nitrogen would have lead to a nitrogen crisis between ~2.2 and possibly as early as

~3.5 Ga (*2, 3*). Biological nitrogen fixation, catalyzed by the nitrogenase enzyme, functions to relieve fixed N limitation (*1, 4*). Today, biological nitrogen fixation is catalyzed by at least three genetically distinct but evolutionarily related nitrogenase enzymes, the most common of which contains Mo (Mo-nitrogenase) while the alternative forms containing V and Fe (V-nitrogenase) or Fe only (Fe-nitrogenase) are less common. The ability to catalyze biological nitrogen fixation is limited to a small subset of phylogenetically distinct bacteria and some of the methanogenic archaea (*5, 6*). Isotopic evidence of $^{15}N/^{14}N$ ratios indicative of biological nitrogen fixation in organic-rich cherts and shales date to >2.5 Ga and suggest an early emergence of the ability to catalyze the reduction of dinitrogen to ammonia (*7-9*).

A nearly complete biological nitrogen cycle is thought to have been present in redox stratified oceans during the late Archean and early Proterozoic (~2.5 to 2.0 Ga) (*8, 10, 11*). During this time, oceans were depleted in Mo and high in Fe, as evidenced by chemical stratigraphic measurements which indicate ancient oceans were limited in soluble Mo, due to the insolubility of Mo-sulfides under anoxic conditions, prior to the rise of oxygen ~2.5 Ga (*12*). The evidence indicating a full biological nitrogen cycle during this time as well as limited soluble Mo lead to the hypothesis that the alternative forms, V-dependent and Fe-nitrogenase predate Mo-nitrogenase and represent the primitive forms of the enzyme (*13, 14*). Recent phylogenetic and evolutionary analysis has lead to an alternative hypothesis for the evolution of biological nitrogen fixation.

Biological Nitrogen Fixation

Nitrogenase is an oxygen sensitive metalloenzyme complex that is considered to be one of the most complex biological innovations. The enzyme has been studied for decades with the first investigations published in the 1960s; however, the extreme oxygen sensitivity of the enzyme and the complex biosynthetic machinery required to assemble an active enzyme has made biochemical characterization challenging. Due to the integral role of biological nitrogen fixation to the global nitrogen cycle and the complexity of the enzyme, the enzyme has become a model system for studying a number of general biological processes such as electron transfer, nucleotide dependent signal transduction and metal-cofactor assembly.

Diazotrophs are organisms capable of catalyzing nitrogen fixation. While diazotrophy is commonly associated with legume symbionts such as *Rhizobium* spp. and *Bradyrhizobium* spp., the phylogenetic and metabolic diversity of bacteria capable of diazotrophy include obligate aerobes and obligate anaerobes as well as facultative anaerobes and microaerophiles and both oxygenic and anoxygenic phototrophs. The process is also found in the archaeal methanogenic orders *Menthanococcales* and *Methansarcinales*, both strict anaerobes (*6, 15*). Due to this diverse array of diazotrophs observed to date, it is not surprising that diazotrophy has been found in a variety of environments including soil, marine and freshwater environments, deep-sea hydrothermal vents and terrestrial hot springs (*16-20*).

The majority of biological nitrogen fixation observed today is catalyzed by the Mo-dependent form of the nitrogenase enzyme (Mo-nitrogenase or *nif*) *(21)*. In fact, the

alternative forms of the enzyme, containing either vanadium and iron (V-nitrogenase or *vnf*) or iron only (Fe-only nitrogenase or *anf*) in the active site are found in a limited subset of diazotrophs that also encode a Mo-nitrogenase (*6, 22, 23*). These enzymes were discovered after cultures with the Mo-dependent form of the enzyme were starved of Mo and fixed nitrogen but were still capable of growing diazotrophically (*24-27*). The alternative forms of the enzyme have not been as well-characterized biochemically as the Mo-nitrogenase and their distribution in nature is much more limited. Gene sequences for the alternative nitrogenases (*anf*/*vnf*) are present in several methanosarcina strains. Vnf regulons are constrained to several proteobacterial and firmicutes genomes as well as one cyanobacterial genome, *Anabaena variabilis*. Anf regulons are also observed in several proteobacterial and firmicutes genomes as well as the genome of *A. variabilis* and a single chlorobi genome. Genetic and growth rate experiments indicate the Mo-nitrogenase is preferred for diazotrophic growth, even when low levels of Mo are present. In addition to the Mo-, V-, and Fe-only nitrogenases, several nitrogenase homologues have been identified in the genomes of several *Firmicutes* and *Chloroflexi*. The active site metal content of these homologues is not known and these organisms have not been shown to grow diazotrophically (*23*). However, several strains of methanogens with nitrogenases with uncharacterized active site metal content have been shown the catalyze nitrogen fixation (*28, 29*).

The nitrogenase enzyme complex consists of two separable iron-sulfur cluster containing proteins that have no measureable enzymatic activity on their own (Fig. 1.1). Together, the protein components act in concert to catalyze the multiple electron reduction of dinitrogen gas to ammonia (*21, 30-36*). All forms of the nitrogenase enzyme

characterized to date are extremely oxygen sensitive and all biochemical analyses of the mechanism and activity of the enzyme require strict anaerobic conditions (*24, 25, 37-40*), despite the observation that diazotrophy is observed in aerobic prokaryotes. Several strategies have been observed in aerobic diazotrophs to protect the nitrogenase enzyme from oxygen inactivation.

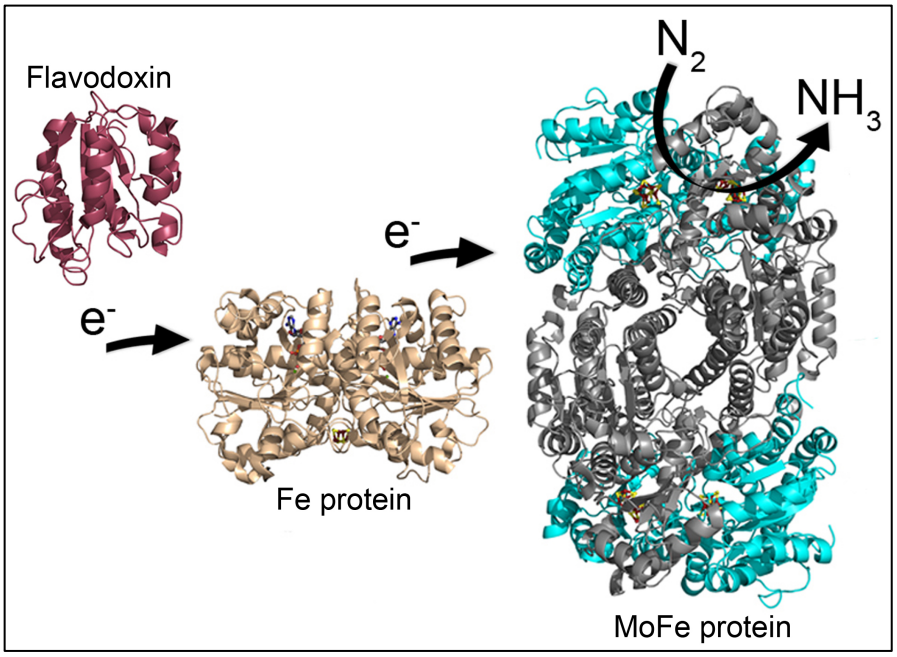


Figure 1.1. Ribbons representation of NifF, a *nif*-specific flavodoxin (raspberry), the Fe Protein (wheat), and the MoFe protein (NifD, cyan, NifK, gray) from *A. vinelandii* with the flow of electrons shown from flavodxoin to the Fe protein and from the Fe protein to the MoFe protein where substrate reduction occurs.

For example, *A. vinelandii*, an aerobic soil bacterium, uses respiratory protection through active consumption oxygen by a specific *bd*-type cytochrome oxidase during diazotorphic growth to protect the nitrogenase enzyme (*41*). In addition, in *Azotobacter vinelandii* and some other diazotrophs, an iron-sulfur protein termed the Shethna protein provides conformational protection of nitrogenase (*42, 43*). For symbiotic nitrogen fixing

organisms, plant nodules on the roots of leguminous plants provides a nearly anaerobic environment where oxygen is kept at very low levels through the activity of leghaemoglobin to support the growth of these microaerophilic diazotrophs (*44, 45*). Many filamentous cyanobacteria form differentiated cells called heterocysts that are anaerobic compartments. Heterocysts are fed carbon by vegetative cells for the energetic needs of nitrogen fixation and in turn, provide vegetative cells with fixed nitrogen (*46*). Temporal separation is utilized by facultative anaerobes to decouple aerobic respiration from nitrogen fixation through regulation in response to oxygen (*47*). For unicellular cyanobacteria present in dense biomass microbial communities, nitrogen fixation is temporally separated from oxygenic photosynthesis, occurring at night when respiration has rendered the mat anaerobic (*48*).

The Nitrogenase Enzyme

Free-living diazotrophs have served as model organisms for studying biological nitrogen fixation due to the ease in culturing and large-scale experiments as opposed to symbiotic nitrogen fixers. To date, the majority of the genetic and biochemical characterization have been aimed at the Mo-nitrogenase in *A. vinelandii*, *Clostridium pasteurianum* and *Klebsiella pneuamoniae*. As a result of initial biochemical studies, two separable iron-sulfur proteins were discovered to comprise the Mo-nitrogenase. The Fe protein (NifH), encoded by *nifH*, is a homodimer bridged by a [4Fe-4S] cluster that serves as the obligate electron donor to the MoFe protein (NifDK), encoded by *nifDK*, a $\alpha_2\beta_2$ heterotetramer containing two complex metalloclusters, the P cluster and the FeMo-cofactor (*49-51*).

The Fe protein associates and dissociates with the MoFe protein for each single electron transfer, coupling each electron transfer event with the hydrolysis of MgATP (*52*) (Figure 1.2).

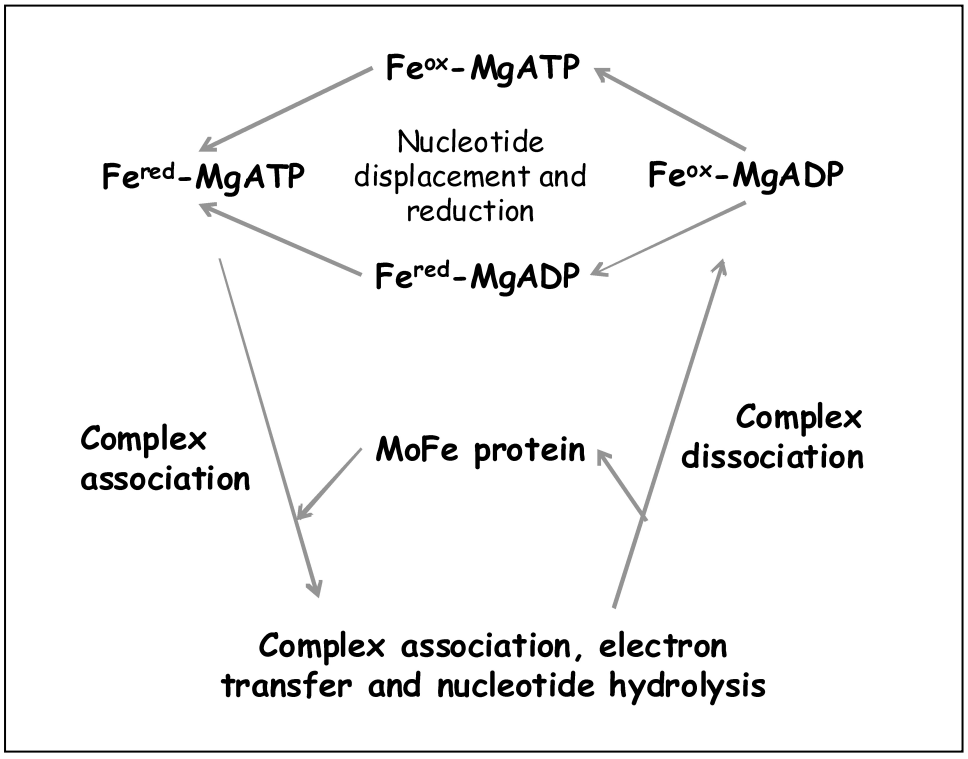


Figure 1.2. The three state cycle of a single nitrogenase Fe protein – MoFe protein complex association and dissociation and electron transfer. Fe protein in the MgADP bound oxidized state (Fe^{ox}-MgADP – upper right) is reduced and undergoes a nucleotide exchange in which MgADP is replaced by MgATP (Fe^{red}-MgATP – upper left). The reduced MgATP bound Fe protein (Fe^{red}-MgATP) then associates with the MoFe protein to form a complex triggering MgATP hydrolysis, electron transfer, and finally dissociation of the oxidized MgADP bound state of the Fe protein (Fe^{ox}-MgADP).

The reduction of dinitrogen to ammonia requires at least 8 electrons per molecule of N_2, which results in the reduction of dinitrogen to ammonia with the concomitant evolution of hydrogen from protons, indicating this association and dissociation must occur at least 8 times. In fact, the dissociation of the protein complexes

is the rate-limiting step in the reduction of dinitrogen (*53, 54*). At least two molecules of MgATP are required for each electron transfer event, making nitrogen fixation very energetically costly. Electrons are shuttled from the Fe protein to the P cluster, a [8Fe-7S] cluster located between the α (NifD) and β subunits of the MoFe protein (NifDK) and then to the FeMo-cofactor, a [Mo-7Fe-8S] cluster in the α subunit where substrate reduction occurs (Fig. 1.1).

The Fe Protein

The Fe protein is responsible for transferring electrons derived from central metabolism in the form of reduced ferredoxin or flavodoxin to the MoFe protein during catalysis (*21, 33, 55*) (Figure 1.1). The Fe protein is the only electron donor to the MoFe protein and requires MgATP hydrolysis for intramolecular electron transfer. This requirement places the Fe protein in a large family of proteins that couple nucleotide binding and hydrolysis to protein conformational change (*56, 57*). However, the Fe protein appears to be a unique member of this class of nucleotide-dependent signal transduction proteins in that most members do not require ATP and do not participate in redox reactions. The Fe protein is a homodimer bridged by a single [4Fe-4S] cluster near the surface of the enzyme (*50, 51, 58*) (Fig. 1.3A, 1.3B). The MgATP binding site is located ~15 Å away from the [4Fe-4S] cluster. The [4Fe-4S] cluster is symmetrically coordinated by four conserved Cys residues, two from each subunit of the homodimer (Fig. 1.3C).

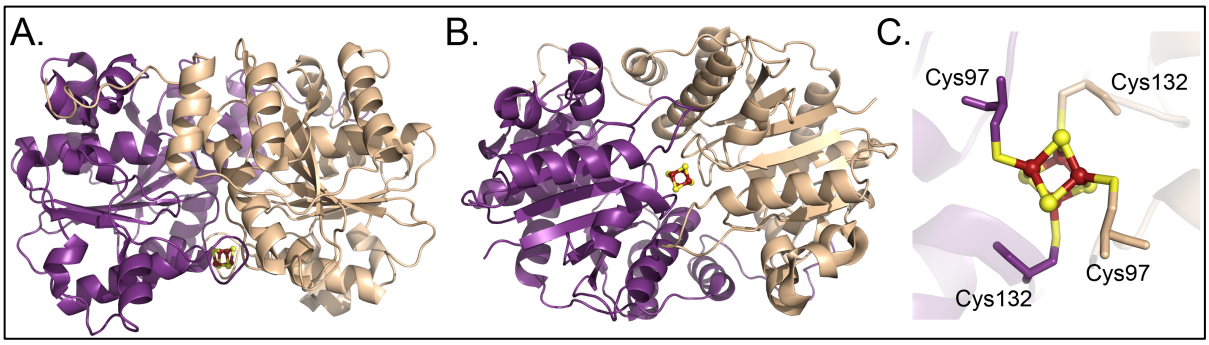


Figure 1.3. Ribbon representation of the overall fold of the *A. vinelandii* Fe protein dimer viewed from two orientations (A and B) down the two fold axis of symmetry. C. An expanded view of the cluster symmetrically coordinated with two analogous Cys ligands from each subunit (amino acid residue numbers are from *A. vinelandii* Fe protein). S, yellow, Fe, firebrick.

This symmetric bridging cluster is rare, observed to date in only two other proteins, 2-hydroxygluraryl-CoA dehydratase and the L protein of protochlorophyllide reductase (*59, 60*). The latter protein, BchL, is a protein component of dark protochlorophyllide reductase (DPOR) involved in chlorophyll biosynthesis. BchL is evolutionarily related to the Fe protein (*61*) and the structure of the Fe protein is similar to BchL (*60*). Despite the unique nature of the bridging cluster, [4Fe-4S] clusters are nearly ubiquitous in biology. These clusters are commonly involved in electron transfer reactions and their oxidation-reduction potentials are sensitive to the surrounding protein environment. The hydrolysis of MgATP only occurs in the presence of the MoFe protein, and is presumed to occur only when the two proteins are in complex. The primary sequence of the Fe proteeals nucleotide-binding motifs that are conserved in other members of the nucleotide-binding protein family (*56, 57, 62*). These motifs are termed the Walker motifs, the Walker A motif (Gly-X-X-X-X-Gly-Lys/Ser) and the Walker B motif (Asp-X-X-Gly), and are responsible for coordinating Mg^{2+} and the phosphate groups (*62, 63*). In addition, key

residues within these motifs are most likely involved in coordinating the γ-phosphate of the nucleoside triphosphate and Mg^{2+}. The phosphate binding loop, or P loop (Gly-X-X-Gly-X-Gly) (Fig. 1.4), adjacent to the Walker A motif, is also conserved in this class of proteins and provides direct hydrogen bonding interactions with the nucleotide phosphate groups.

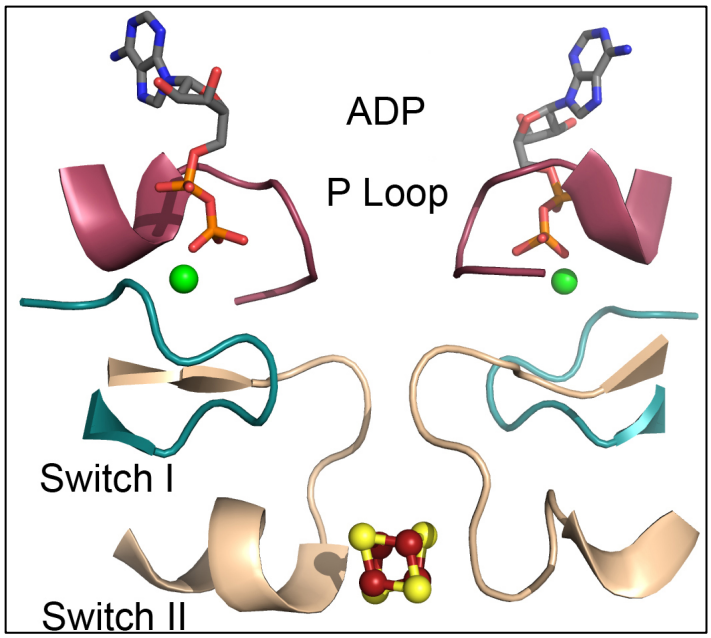


Figure 1.4. Ribbon representation of the two nucleotide dependent Switch regions (Switch I, teal, Switch II, wheat) and the phosphate binding 'P loop' region (raspeberry) of the Fe protein shown relative to the location of the nucleotide (stick representation) in the structure of the MgADP state of the Fe protein. S, yellow, Fe, firebrick, O, red, P, orange, N, blue, C, gray, Mg, green.

Amino acids of the switch regions I and II provide pathways for communication between nucleotide binding and the site for MoFe protein (Switch I) and the [4Fe-4S] cluster (Switch II) (*64-66*) (Figure 1.4). The Switch regions detect changes in the nucleotide bound states of the Fe protein, coupling the protein conformational changes in the Fe

protein to signal MoFe protein docking and complex formation or to impose changes on the [4Fe-4S] cluster (*21, 67-75*).

The MoFe Protein

The MoFe protein is a $\alpha_2\beta_2$ heterotetramer. The α-subunit (NifD) is encoded by *nifD* while the β-subunit (NifK) is encoded by *nifK*. The two proteins share significant primary sequence identity and are the product of an ancient gene duplication event (*23, 76*). The crystal structure of the MoFe protein from *A. vinelandii, C. pasteurianum* and *K. pneumoniae* reflects the primary sequence similarity in NifD and NifK, resulting in overall folds that are similar (*77-83*). The α- and β-subunits are similar in size, each existing as three domains of a single parallel β-sheet and α-helices with each αβ dimer serving as a symmetric half of the MoFe protein (Fig. 1.5).

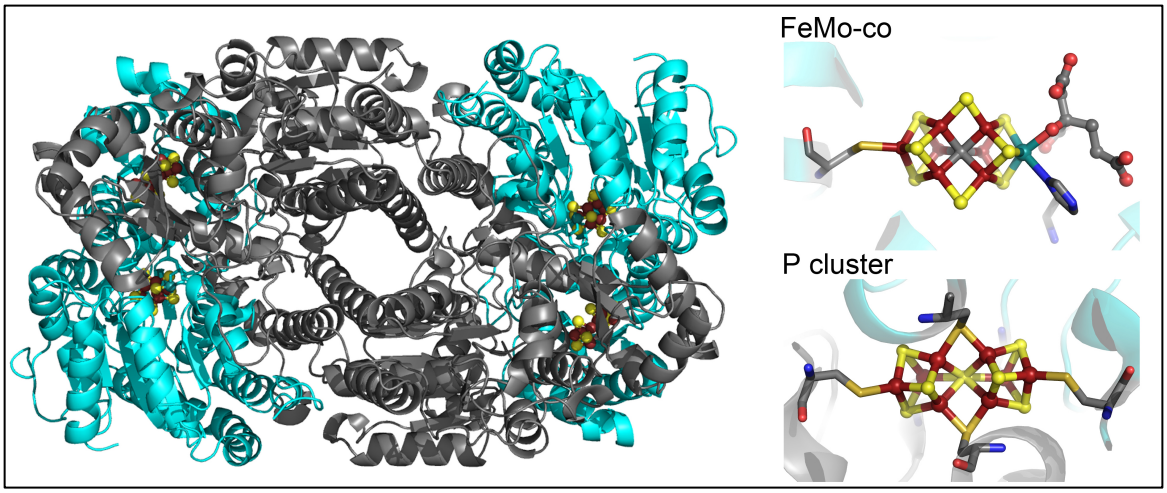


Figure 1.5. A. Ribbons representation of the nitrogen MoFe protein from *A. vinelandii* viewed down the twofold axis of symmetry that relates each equivalent dimer of the $\alpha_2\beta_2$ heterotetramer. Two FeMo-cofactors (B) are located solely within the two α- subunits. Each α- and β-subunits are similar in structure and are related by pseudo-twofold symmetry. The MoFe protein P clusters (C) are coordinated symmetrically by three Cys ligands of each subunit in a similar fashion to the [4Fe–4S] cluster of the Fe protein which is coordinated symmetrically by equivalent Cys residues of two separate protein chains. S, yellow, Fe, firebrick, N, blue, O red, Mo, deep teal.

MoFe protein, similar to the Fe protein, is evolutionarily related to proteins of the DPOR complex, BchNB (*61*). These proteins are also part of a multi-protein complex (with BchL) that catalyze the double bond reduction of protochlorophyllide (*60, 84-90*). Unlike the MoFe protein, BchNB contains only a [4Fe4S] cluster at the analogous P cluster site and retains and a cavity for substrate binding where FeMo-co binds in the nitrogenase enzyme (*84, 87*).

The P cluster, which lies at the interface of the α- and β-subunit, serves to mediate electron transfer from the [4Fe-4S] of the Fe protein to the FeMo-co. Structures of the nitrogenase complex show the P cluster lies between the Fe protein [4Fe-4S] and the FeMo-co in NifDK (*77, 78, 81-83, 91, 92*) and the distance between the [4Fe-4S] of the Fe protein and FeMo-co is too far for typical electron transfers (*64-66, 93*). In addition, amino acid substitutions between the P cluster (*83*) and FeMo-co perturb intramolecular electron transfer. Also, spectroscopic studies indicate the oxidized P cluster can be reduced by the Fe protein (*94*) which provides evidence that the P cluster mediates electron transfer from the Fe protein to the active site metal cluster for substrate reduction. The P cluster is a rare [8Fe-7S] cluster in which the uneven ratio of Fe:S is overcome by a terminal bridging thiolate (*83*) (Fig. 1.6A). FeMo-co, the metal cofactor where dinitrogen reduction occurs, is housed in the α-subunit of the MoFe protein. The cofactor is one of the most complex in biology and is described as a complex bridged metal assembly consisting of two modified [4Fe-4S] cluster cubanes in which one cluster contains a Mo substituted at one of the Fe positions (Fig. 1.6B).

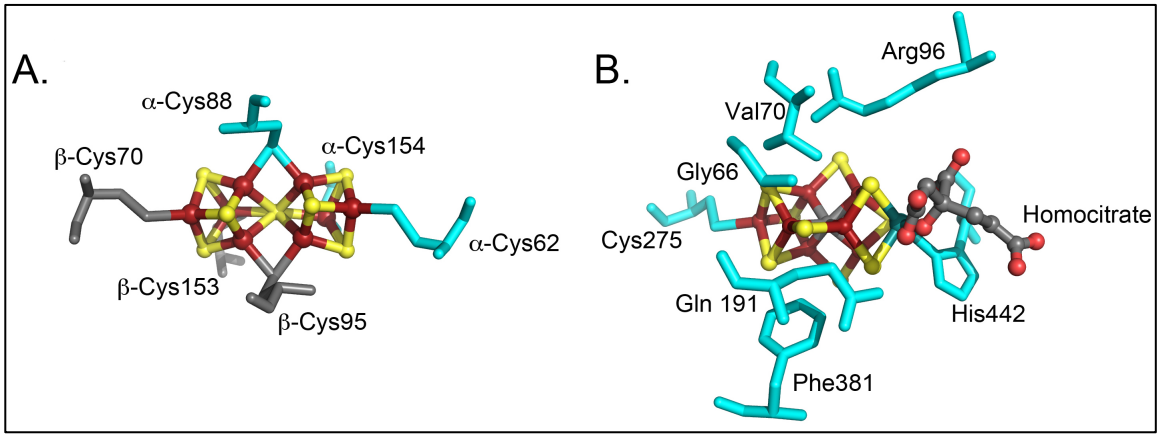


Figure 1.6. P cluster (A) and FeMo-co binding sites (B) in the MoFe protein structure with ligands and conserved residues of the FeMo-co binding site indicated. Residues from the α-subunit (NifD) are colored cyan and residues of the β-subunit (NifK) are colored gray. Residue numbers correspond to the amino acid sequence of NifD and NifK from *A. vinelandii.* Fe, firebrick, S, yellow, C, gray, Mo, teal.

Protein coordination of the FeMo-co is very limited with a single Cys ligand at the terminal Fe position and a single His coordination at the Mo. Despite this limited coordination, the residues of the binding site pocket are well conserved. Key mutagenesis studies pre-dating the structural characterization of the MoFe protein indicated the coordinating ligands residues as well as residues of the cluster environment were important for nitrogenase function (*95-100*).

Biosynthesis of Nitrogenase

Biosynthesis and maturation of the nitrogenase enzyme requires a suite of *nif*-regulated genes encoding various scaffolds, maturases, and carrier proteins, most of which are necessary for synthesizing FeMo-co (*1, 101*) (Fig. 1.7).

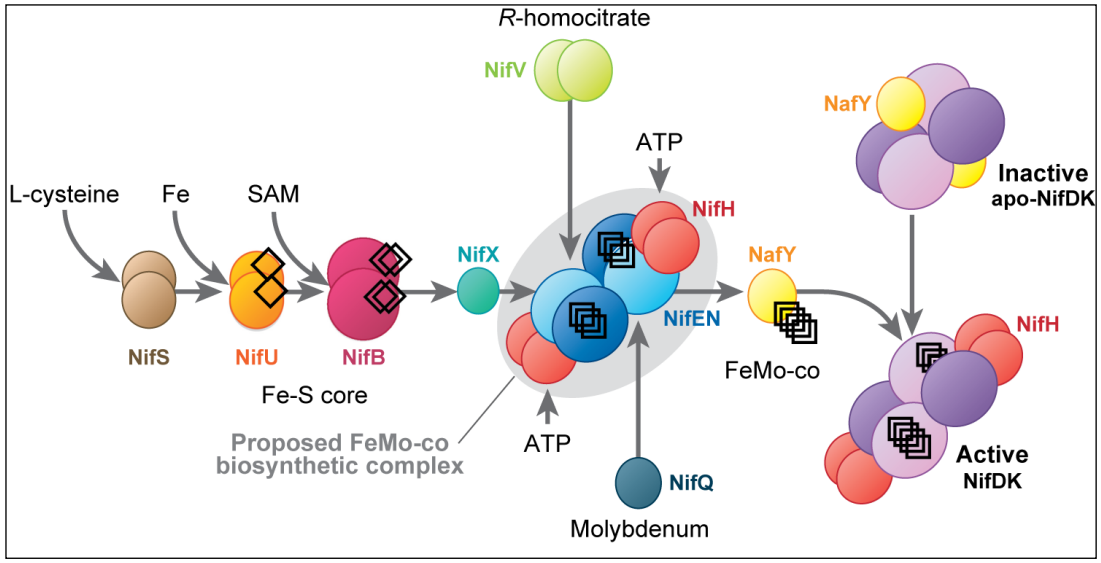


Figure 1.7. Model for the biosynthesis of FeMo-co. Scheme illustrating the convergence of FeMo-co 'building blocks' into a central node composed of the NifEN and the Fe protein. The metallocluster carrier proteins, NifX and NafY, serve as links between the Fe–S core biosynthetic branch, the NifEN/NifH central node, and the apo-MoFe protein, respectively (from reference (*1*)).

Nitrogenase maturation has been an intriguing biochemical question since the first analyses of biological nitrogen fixation owing to the complexity of the enzyme and the metal clusters necessary to achieve catalysis and rivals the complexity of the enzyme itself. The majority of the characterization of the biosynthietic pathway has been carried out in the model organisms *K. pneumoniae* and *A. vinelandii* using a variety of techniques including gene deletions and complementation (*102-105*) and through the development of cell-free assays (*106*). NifD and NifK, the protein components of NifDK, are not required for the synthesis FeMo-co (*107*). Instead, FeMo-co is assembled elsewhere and inserted in the apo-MoFe protein to generate the active holo-protein. The P clusters must be present in the MoFe protein for FeMo-co insertion to occur (*108*). The biosynthesis of FeMo-co and thus a holo-MoFe protein begins with simple Fe-S clusters. These are

assembled by the activity of NifU and NifS, homologues of the Isc (iron-sulfur cluster) housekeeping FeS cluster machinery, that are *nif*-regulated and often recruited to the *nif*-cluster for the purpose of FeS cluster biosynthesis necessary to sustain biological nitrogen fixation (*109-112*). NifS is a cysteine desulfurase, providing the S necessary for cluster assembly while NifU serves as a molecular scaffold for cluster assembly, transferring clusters to apo-proteins by direct protein-protein interactions. NifU is responsible for transferring a [4Fe-4S] cluster to the Fe protein and NifU and NifS may also be involved in P cluster biosynthesis, though its direct role is not known. The presence of P clusters in the apo-MoFe protein result in conformational changes, bringing the α- and β- subunits together and resulting in an open FeMo-co binding site (*113, 114*).

The biosynthesis of FeMo-co involves the scaffold proteins NifU, NifB, NifQ, and NifEN, the carrier proteins NifX and NafY and the enzymes NifS, NifB, NifV and NifH (*102, 103, 115-117*) (Fig. 1.7). Although the maturation of the FeMo-co requires the Fe protein, the function of the Fe protein in the biosynthetic pathway is not known. Building blocks of FeMo-co are assembled on the molecular scaffold NifEN (*95*), a protein that shares sequence and structural similarity with the MoFe protein (*23, 76, 118, 119*) (Fig. 1.8). Like NifD and NifK, NifE and NifN also arose from an ancient gene duplication event wherein NifDK duplicated to give rise to the scaffold protein NifEN (*76, 118*). The Fe-S core of the FeMo-co is assembled by NifU, NifS and NifB. Simple Fe-S clusters are provided to NifB by NifU and NifS. NifB constains a S-adenosylmethionine (SAM) radical domain at the N-terminus (*120*) and synthesizes NifB-co, a more complex Fe-S cluster, in a reaction dependent on radical SAM chemistry (*121, 122*).

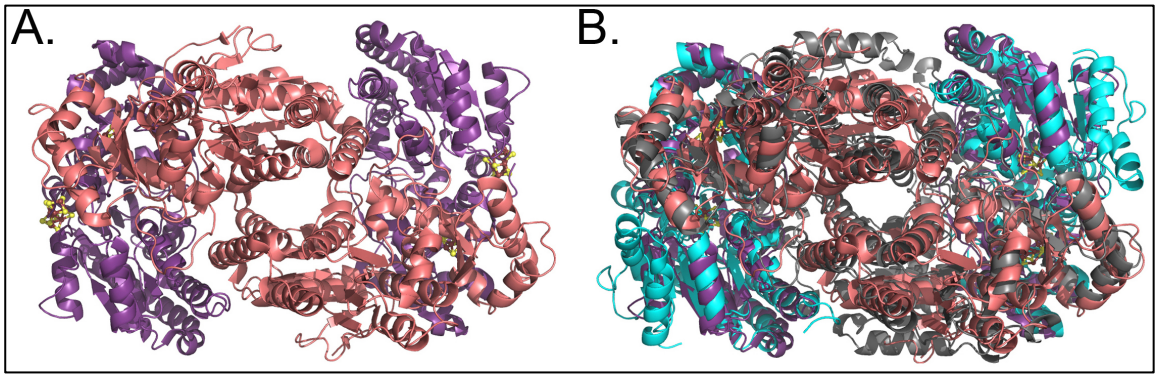


Figure 1.8. Ribbons representation of NifEN from *A. vinelandii* (A) and the NifEN protein superimposed on the MoFe protein from *A. vinelandii* (B) indicating the structural similarity between NifEN and the MoFe protein. NifE is shown violet, NifN is shown in salmon, NifD is shown cyan, and NifK is shown in gray.

NifB-co is then shuttled to NifEN on a carrier protein, such as NifX or transferred directly from NifB. NifQ, an iron-sulfur protein, is involved in Mo acquisition and serves as the Mo donor to FeMo-co (*123*). Homocitrate, the non-protein ligand that coordinates the Mo of FeMo-co (Fig. 1.6A), is synthesized by homocitrate synthase, and like Mo, is incorporated into the immature FeMo-co on the scaffold NifEN (*124*). The mature FeMo-co is then transferred directly to apo-MoFe or delivered by a carrier protein, resulting in the formation of a holo-MoFe protein capable of dinitrogen reduction (*1*).

The biosynthetic machinery necessary for assembling an active nitrogenase requires a suite of *nif*-encoded genes in addition to the presence of the structural genes *nifH*, *D* and *K*; however, the recruitment and co-localization of these genes varies among diazotrophs. The structural genes are usually co-located within an operon in the order *H-D-K* and are in close proximity to or in the same operon of *nifE* and *nifN* genes. Many organisms, including *A.vinelandii*, contain an extensive collection of genes involved in nitrogen fixation and regulation (*102, 125, 126*) (Fig. 1.9A). However, the nitrogenase genetic machinery in diazotrophic Euryarchaeota is simpler, where five of the six gene

products necessary for nitrogen fixation, *nifH*, *nifD*, *nifK*, *nifE*, and *nifN* are co-located in one operon and *nifB* is located downstream (*127*) (Fig. 1.9B).

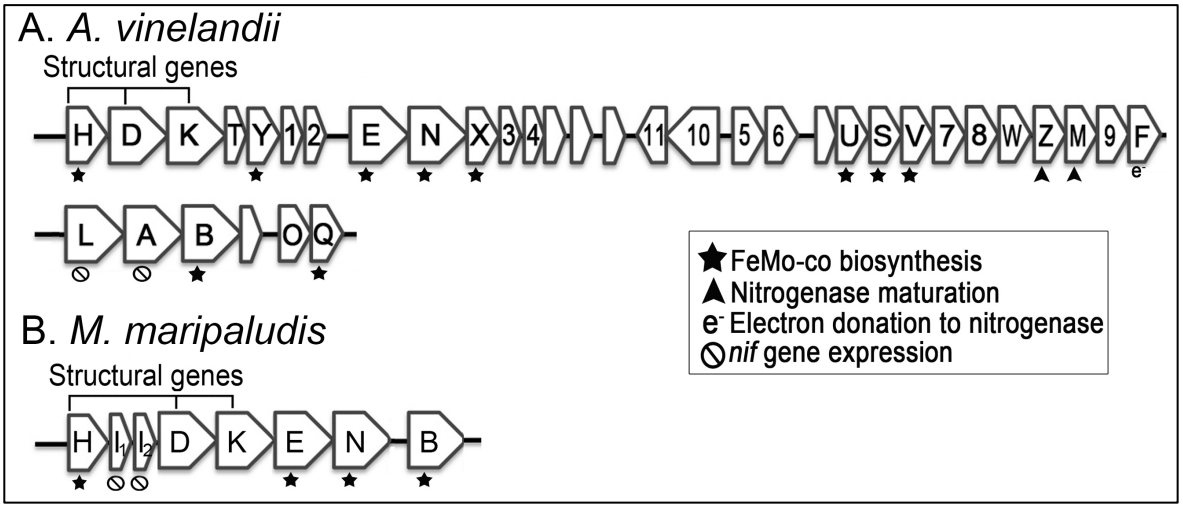


Figure 1.9. Organization of nitrogen fixation (*nif*) genes in *A. vinelandii* and *Methanococcus maripaludis* (B). Shown are all those genes given the designation *nif* and associated open reading frames that are part of *nif* specific transcriptional units.

A recently developed bacterial *in vitro* system containing proteins NifEN, NifB and the Fe protein (NifH) as well as homocitrate, MgATP and S-adenosyl methionine is sufficient to assemble a complete FeMo-co and deliver it to the MoFe (NifDK) protein (*105*). This expendability of accessory gene products in some diazotrophs *in vivo* suggests the assembly of the FeMo-cofactor may proceed without these proteins or that other assembly and scaffold proteins, possibly from the ISC or SUF iron sulfur cluster assembly systems, are able to substitute in the biosynthetic pathway in some prokaryotes.

Distribution of Biological Nitrogen Fixation

Nitrogen fixation occurs under a variety of physical and geochemical conditions;

however, the influence of these factors on the distribution, diversity, and activity of organisms capable of fixing nitrogen have yet to be defined. Metal availability, ammonia levels, and ample energy supply are all factors that contribute to the occurrence of nitrogenase activity. Each of these factors is influenced by pH and temperature both of which presumably constrain nitrogenase activity at some threshold level. Molybdenum is rarely a limiting factor in natural environments, but must be present for Mo-dependent nitrogenase biosynthesis to occur. In rare cases microorganisms maintaining alternative nitrogenases will fix nitrogen if molybdenum is not available. If V is available, then organisms encoding Vnf regulons will grow diazotrophically throught the activity V-nitrogenase, and if neither Mo nor V are available, organisms that encode an Fe-only-nitrogenase enzyme will grow diazotrophically using this enzyme (*128*). The nitrogenase enzyme complex is extremely oxygen-sensitive and is regulated at the transcriptional level when excess oxygen is present (*129*), and as a result, the strategies mentioned above are employed by diverse diazotrophs to protect the nitrogenase enzyme from oxygen inactivation. The upper temperature constraint of nitrogen fixation is still being explored. There are several reports of diazotrophy at ~64°C, and one observation of nitrogen fixation by an organism isolated from a hydrothermal sea vent at temperatures up to 92°C (*29, 130, 131*).

Yellowstone National Park— Abundant Geochemical Gradients

The Earth formed approximately 4.5 billion years ago, and indications for the emergence of life have been dated back 3.8 Gyr. If this is true, then life persisted on Earth

for 1.5 billion years before the emergence of oxygen (*132-135*). Geologic and geochemical evidence to date suggests that the first organisms on Earth originated and survived under anaerobic conditions using chemotrophic metabolisms, obtaining energy from reduced chemical species generated through electrical discharge, meteoric impacts, and/or volcanic sources (*134-136*). Concurrent phylogenetic analysis indicates that the earliest branching lineages are comprised of thermophilic and hyperthermophilic chemoheterotrophic and chemolithotrophic metabolisms (*135-137*).

Prior to the emergence of cyanobacteria and the eventual rise of oxygen in the atmosphere, early metabolic regimes likely relied upon electron donors such as H_2, H_2S, or Fe^{2+} in the conversion of CO_2 into biomass (*135, 136, 138-140*). Carbon- and sulfur-isotopic analysis has dated the biogenic production of methane and methanogenic and bacterial sulfate reduction to the Archean period (*138-140*). Evidence from Archean aged crust suggests that smaller crust fragments existed during this time, leading to higher rates of subduction and cycling of these smaller fragments which would have driven increased hydrothermal cycling, and thus the occurrence of hydrothermal environments would have been much more widespread than observed on Earth today (*141*). A hydrothermal environment in the Archean would have supported organisms capable of methanogenesis and subsequent methane production (*134, 136, 138, 140, 142*). The hypothesis that life first emerged in a high temperature environment is supported by phylogenetic analyses wherein the most deeply-rooted organisms are thermophiles suggesting that ancient life was thermophilic (*143-145*). Molecular characterization of phylogenetic markers such as 16sRNA genes in modern thermophilic microbial communities reveal the presence of organisms that branch most basal to the last common

ancestor and might be the closest modern-day relatives of the microbial populations that dominated Earth under Archean conditions when life first emerged (*137, 139, 142, 144*).

Microbial Diversity in Extreme Environments of Yellowstone National Park

The physical and chemical gradients present in the geothermal features of Yellowstone National Park (YNP) provide a variety of ecological niches for surveying microbial life and the diversity therein. Hydrothermal activity driven by heat from the underlying magma chamber results in a range of temperatures and chemical composition of surface and subsurface fluids, the chemical makeup of which vary as a result interaction with the rocks they encounter enroute to the Earth's surface, and subsurface processes. Chemical compositions between and within thermal features vary, often ranging in concentrations gradients of several orders of magnitude (*139*). Boiling, circum-neutral springs dominated by chloride and silica are found throughout the park, including thermal features at Heart Lake and Imperial Geyser (Fig. 1.10). There are also examples of acid-sulfate pools such as those found in the Norris Geyser Basin (Fig. 1.10).

Numerous metagenomic and phylogenetic analyses have been conducted in YNP in geothermal features spanning large temperature and geochemical gradients. As a result, genetic and ecological techniques have been used to better understand the microbial diversity and metabolic complexity of a number of ecological niches found within the YNP geothermal complex. Focused studies of hot springs above ~70°C have also been conducted on the basis of available energy sources (*146*). As photosynthesis is not likely to occur at these temperatures, the microbial populations inhabiting these features are predominantly chemolithotrophs, capable of, among many other

metabolisms, oxidizing hydrogen and sulfide for fixing CO_2 (*147*). These microbial populations varied as a result of hydrogen concentration, but the dominant constituents were members of the *Aquaficales (148).* Members of the *Thermotogales, Thermus/Deinococcus,* and *Thermodesulfobacteria* were also detected.

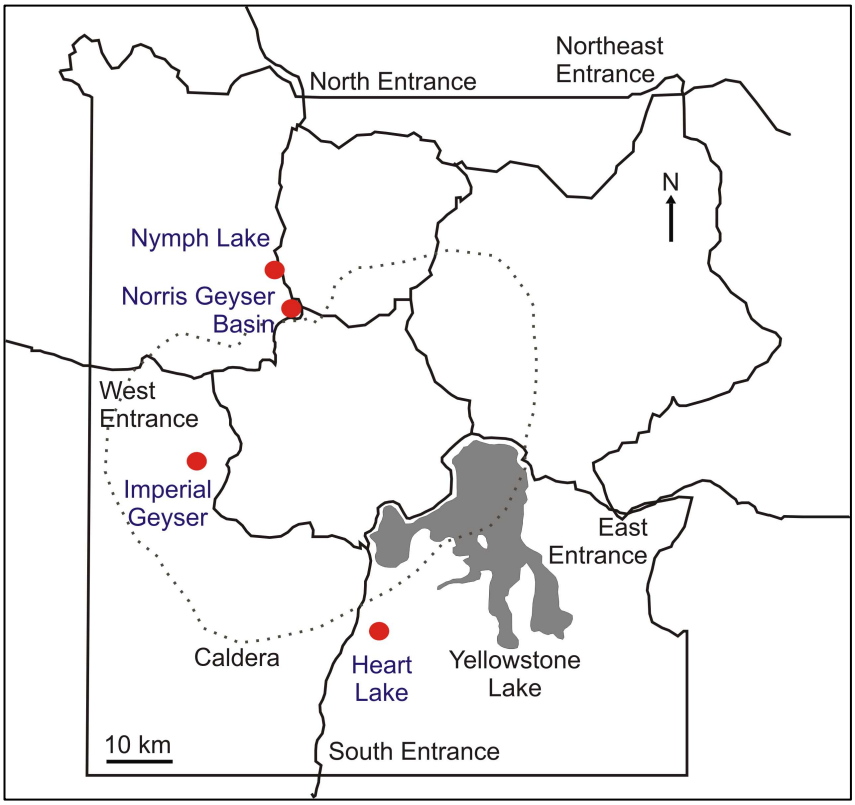


Figure 1.10. Map of Yellowstone National Park with primary sampling sites indicated by red circles.

Contained within YNP are a number of acid-sulfate-chloride (ACS) springs with pH values less than 4. These acidic features vary in temperature and often contain trace elements, such as mercury, boron, and arsenic, at concentrations greater than what is considered to be toxic. Some acid-tolerant algae and cyanobacteria are found in the lower temperature acidic springs (40- 55°C), but most of these ACS springs are dominated by

chemolithotrophs such as members of *Hydrogenobaculum* and *Desuferella (149)*. Natural hydrocarbon seeps such as Rainbow Springs in YNP are home to predominantly heterotrophic acidophilic bacteria; however, these springs also host iron- and sulfur-oxidizing chemolithotrophic bacteria (*150*). These acidic springs have been shown to contain elevated concentrations of sulfur compounds.

Microbial mat communities are found throughout YNP and house a variety of phototrophic and nitrogen fixing microbial populations, including both oxygenic and anoxygenic phototrophs (*151*). For example, 16S rRNA analysis of mats harvested from alkaline silica-depositing hot springs, namely Octopus Spring, contain a top layer of phototrophs, consisting of unicellular *Synechococcus* spp. and filamentous green nonsulfur bacteria such as *Roseiflexus* spp. Anoxygenic phototrophs, such as *Chlorobaculum tepidum,* a green sulfur bacterium, and a purple bacterium, *Thermochromatium tepidum* have been found at temperatures near 50°C in deeper portions of the sections where oxygen is less abundant. Methanogens and chemoorganotrophs have also been found in these classes of proteobacteria; however, these organisms are constrained to the anaerobic portions of the mat community.

Members from both the Euryarchaeota and Crenarchaeota have been detected in thermal features of YNP, and in fact some of the first Archaea were found in these environments (*152*). Novel groups of Archaea were discovered by phylogenetic analysis of ribosomal RNA sequences from Obsidian Pool, with temperatures ranging from 75°C to 95°C. Several of these sequences diverged below the bifurcation of between Euryarchaeota and Crenarchaeota, resulting in a new phylum of Archaea, the Koarchaeaota that branches more basal to LUCA than the Crenarchaeota (*152, 153*).

Archaea are often the dominant populations found in extreme environments and may represent the only microbial organisms capable of inhabiting some of the most extreme environments. The thermal features found at Heart Lake, Nymph Lake, Imperial Geyser, and Norris Geyser Basin provide geochemical and physical gradients necessary for examining the environmental factors that constrain the occurrence of biologically-mediated processes such as diazotrophy. The acid thermal waters of Norris Geyser Basin are some of the hottest in the Park. Imperial Geyser and Heart Lake also feature alkaline thermal springs while thermal features of the Nymph Lake area are predominantly acidic.

Research Directions

Over the past decade, the advent of whole genome sequencing and metagenome sequencing has vastly increased the amount of microbial genetic data. This increase in available data has enabled the examination of a number of evolutionary questions, including the origin and evolution of biological nitrogen fixation. Using genetic events such as gene duplication and gene fusion events of the structural and biosynthetic genes necessary for the catalysis of dinitrogen reduction, key insight has been gained in elucidating the path of nitrogenase emergence and evolution. In addition, this increase in sequence data has provided the impetus for examining the occurrence of essential microbial processes such as biological nitrogen fixation in natural environments. These analyses provide details regarding the limits of these processes on Earth today as well as physical and geochemical parameters governing the evolution of enzymes necessary for the perpetuation of life on Earth.

The advent of biological nitrogen fixation most likely had a profound impact on the evolution of life on Earth. Nitrogenase is a multicompenent enzyme relying on several complex metalloclusters and a suite of genes to synthesize an active enzyme. Key phylogenetic events such as gene fusions and gene duplications of the biosynthetic and structural genetic machinery yield insight into the origin and evolution of the nitrogenase enzyme. Information gleaned from biochemical and genetic information has indicated the core set of gene products necessary to catalyze the reduction of dinitrogen. In chapter 2, phylogenetic examination of these key genetic events in the evolution of the nitrogenase enzyme should reveal when biological nitrogen fixation emerged and in what sort of organism, and possibly in what geochemical environment.

The ability to catalyze nitrogen fixation is distributed throughout phylogenetically diverse prokaryotes, including obligate aerobes and obligate anaerobes as well as facultative anaerobes. In addition, diazotrophs are found in a number of natural environments ranging from deep sea hydrothermal vents to the soil rhizosphere. This widespread distribution in nature and the observed phylogenetic diversity of known diazotrophs provides little evidence into the ecological constraints that govern the occurrence of biological nitrogen fixation. Since the ability to fix nitrogen was presumably integral to the evolution of life on Earth as abiotic sources of fixed N became limiting, examining the limits of diazotrophy in natural environments present on Earth today could yield important information regarding the environment and thus the metabolic characteristics in which the nitrogenase enzyme arose. The highly conserved nature of the nitrogenase enzyme has enabled the design and implementation of degenerate PCR primers to amplify functional genes, specifically *nifH*, to examine the

distribution and diversity of potential diazotrophs in the thermal features of YNP, where geochemical and spatial gradients are abundant. In chapter 3, degenerate primers were designed to capture the known diversity of *nifH*, facilitating a PCR-based examination of ecological parameters that underpin the distribution and diversity of diazotrophs in these extreme environments in YNP. The distribution of putative diazotrophs in YNP as well as other natural environments is widespread. One aspect for which little is known is the occurrence of putative diazotrophs in high temperature, low pH environments, a question that can be answered through sampling of acidic hydrothermal systems in YNP. In chapter 4, partial enrichment, gene abundance data and temperature- and pH-responsive activity assays indicate diazotrophic populations are wide-spread in acidic, high temperature springs in YNP and these populations are adapted to local environmental conditions.

Phylogenetic, biochemical and evolutionary examination of diazotrophy has revealed a complex set of biosynthetic machinery necessary for assembling an active nitrogenase enzyme. The discovery and subsequent limited characterization of the alternative forms of the enzyme reveals that these systems lack the required genes encoding biosynthetic machinery necessary for assembly of their respective active site cofactors. This lack of biosynthetic machinery and the observation that not all diazotrophs encode alternative nitrogenase regulons raises important questions regarding the environmental conditions, such as Mo limitation, that would have provided the selective pressure resulting in the alternative forms of the enzyme. Furthermore, the observation that Mo-nitrogenase is preferentially expressed even at very low Mo concentrations, suggests that expression of the required biosynthetic machinery for each

form of nitrogenase may be tightly regulated. In chapter 5, decades of biochemical and mutagenesis experiments, phylogenetic analyses and total RNA sequencing were combined to address these interesting aspects of nitrogenase evolution and expression as a result of active site metal content. Transcriptional profiling experiments were carried out in *Azotobacter vinelandii*, a model diazortroph, to determine the response of this organism to diazotrophic conditions with varying metal availabilty. *A. vinelandii* encodes all three forms of the nitrogenase enzyme and has served as a model organism for studying biological nitrogen fixation for decades. The recent completion of the genome of *A. vinelandii* and the advent of full transcriptome sequencing enables a comprehensive analysis of transcript abundance as a result of metal availability in the absence of fixed nitrogen. This study further elucidated the suite of genes necessary for diazotrophic growth in the absence of Mo, providing key insight into the crosstalk between nitrogenase biosynthetic machinery as well as indicating the effects of biological nitrogen fixation on central metabolism in this organism. Chapter 6 details the differential accumulation of *nif* structural gene mRNA, combining transcript abundance data with northern blot hybridization analyses. This study identified intergenic RNA secondary structures that affect the transcription of *nif* structural genes, indicating these secondary structures lead to the high Fe protein to MoFe protein ratio necessary for optimal diazotrophic growth. Chapter seven provides a summary of the major results in each chapter and discusses how these results have greatly improved our understanding of the environmental factors that constrain the distribution and activity of biological nitrogen fixation as well as providing key insight into the crosstalk of the nitrogenase biosynthetic machinery and the integration of biological nitrogen fixation into central metabolism.

Collectively, the results presented in this dissertation enable a greater understanding of the evolution of the nitrogenase enzyme, including the environmental factors favoring the emergence and driving the evolution of the enzyme and providing additional evidence that the Mo-dependent nitrogenase is ancestral to the alternative forms of the enzyme.

Literature Cited

1. Rubio, L. M., and Ludden, P. W. (2008) Biosynthesis of the iron-molybdenum cofactor of nitrogenase, *Annu Rev Microbiol 62*, 93-111.

2. Kasting, J. F., and Siefert, J. L. (2001) Biogeochemistry: The nitrogen fix, *Nature 412*, 26 - 27.

3. Navarro-González, R., McKay, C. P., and Mvondo, D. N. (2001) A possible nitrogen crisis for Archaean life due to reduced nitrogen fixation by lightning, *Nature 412*, 61-64.

4. Zehr, J. P., Waterbury, J. B., Turner, P. J., Montoya, J. P., Omoregie, E., Steward, G. F., Hansen, A., and Karl, D. M. (2001) Unicellular cyanobacteria fix N2 in the subtropical North Pacific Ocean, *Nature 412*, 635-638.

5. Boyd, E. S., Anbar, A. D., Miller, S., Hamilton, T. L., Lavin, M., and Peters, J. W. (2011) A late methanogen origin for molybdenum-dependent nitrogenase, *Geobiol. 9*, 221–232.

6. Raymond, J., Siefert, J. L., Staples, C. R., and Blankenship, R. E. (2004) The natural history of nitrogen fixation, *Mol Biol Evol 21*, 541-554.

7. Beaumont, V., and Robert, F. (1999) Nitrogen isotope ratios of keragens in Precambrian cherts: a record of the evolution of atmospheric chemistry?, *Precambrian Res. 96*, 63-82.

8. Godfrey, L. V., and Falkowski, P. G. (2009) The cycling and redox state of nitrogen in the Archaean ocean, *Nature Geosci 2*, 725-729.

9. Garvin, J., Buick, R., Anbar, A. D., Arnold, G. L., and Kaufman, A. J. (2009) Isotopic evidence for an aerobic nitrogen cycle in the latest Archean, *Science 323*, 1045-1048.

10. Canfield, D. E., Glazer, A. N., and Falkowski, P. G. (2010) The evolution and future of Earth's nitrogen cycle, *Science 330*, 192-196.

11. Garvin, J., Buick, R., Anbar, A. D., Arnold, G. L., and Kaufman, A. J. (2009) Isotopic Evidence for an Aerobic Nitrogen Cycle in the Latest Archean, *Science 323*, 1045 - 1048.

12. Helz, G. R., Miller, C. V., Charnock, J. M., Mosselmans, J. F. W., Pattrick, R. A. D., Garner, C. D., and Vaughan, D. J. (1996) Mechanism of molybdenum removal from the sea and its concentration in black shales: EXAFS evidence, *Geochim. Cosmochim. Acta 60*, 3631-3642.

13. Anbar, A. D., Duan, Y., Lyons, T. W., Arnold, G. L., Kendall, B., Creaser, R. A., Kaufman, A. J., Gordon, G. W., Scott, C., Garvin, J., and Buick, R. (2007) A whiff of oxygen before the great oxidation event?, *Science 317*, 1903-1906.

14. Scott, C., Lyons, T. W., Bekker, A., Shen, Y., Poulton, S. W., Chu, X., and Anbar, A. D. (2008) Tracing stepwise oxygenation of the Proterozoic ocean, *Nature 452*, 456-459.

15. Leigh, J. A. (2000) Nitrogen fixation in methanogens: the archaeal perspective, *Current issues in molecular biology 2*, 125-131.

16. Affourtit, J., Zehr, J. P., and Paerl, H. W. (2001) Distribution of Nitrogen-Fixing Microorganisms along the Neuse River Estuary, North Carolina, *Microb Ecol 41*, 114-123.

17. Mehta, M. P., Butterfield, D. A., and Baross, J. A. (2003) Phylogenetic diversity of nitrogenase (nifH) genes in deep-sea and hydrothermal vent environments of the Juan de Fuca Ridge, *Appl Environ Microbiol 69*, 960-970.

18. Burns, J. A., Zehr, J. P., and Capone, D. G. (2002) Nitrogen-fixing phylotypes of Chesapeake Bay and Neuse River estuary sediments, *Microb Ecol 44*, 336-343.

19. Hamilton, T. L., Boyd, E. S., and Peters, J. W. (2011) Environmental Constraints Underpin the Phylogenetic Diversity of nifH in the Yellowstone Geothermal Complex, *Microb Ecol*.

20. Hall, J. R., Mitchell, K. R., Jackson-Weaver, O., Kooser, A. S., Cron, B. R., Crossey, L. J., and Takacs-Vesbach, C. D. (2008) Molecular characterization of the diversity and distribution of a thermal spring microbial community by using rRNA and metabolic genes, *Appl Environ Microbiol 74*, 4910-4922.

21. Burgess, B. K., and Lowe, D. J. (1996) Mechanism of Molybdenum Nitrogenase, *Chem Rev 96*, 2983-3012.

22. Boyd, E. S., Anbar, A. D., Miller, S., Hamilton, T. L., Lavin, M., and Peters, J. W. (2009) An Early Origin for Molybdenum-Nitrogenase, *Science In review*.

23. Soboh, B., Boyd, E. S., Zhao, D., Peters, J. W., and Rubio, L. M. (2011) Substrate specificity and evolutionary implications of a NifDK enzyme carrying NifB-co at its active site, *FEBS Lett 584*, 1487-1492.

24. Chisnell, J. R., Premakumar, R., and Bishop, P. E. (1988) Purification of a second alternative nitrogenase from a nifHDK deletion strain of Azotobacter vinelandii, *J Bacteriol 170*, 27-33.

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

26. Robson, R. L., Eady, R. R., Richardson, T. H., Miller, R. W., Hawkins, M., and Postgate, J. R. (1986) The alternative nitrogenase of *Azotobacter chroococcum* is a vanadium enzyme, *Nature 322*, 388-390.

27. Schneider, K., Muller, A., Schramm, U., and Klipp, W. (1991) Demonstration of a molybdenum- and vanadium-independent nitrogenase in a nifHDK-deletion mutant of Rhodobacter capsulatus, *European journal of biochemistry / FEBS 195*, 653-661.

28. Dekas, A. E., Poretsky, R. S., and Orphan, V. J. (2009) Deep-sea archaea fix and share nitrogen in methane-consuming microbial consortia, *Science 326*, 422-426.

29. Mehta, M. P., and Baross, J. A. (2006) Nitrogen fixation at 92 degrees C by a hydrothermal vent archaeon, *Science 314*, 1783-1786.

30. Hoffman, B. M., Dean, D. R., and Seefeldt, L. C. (2009) Climbing nitrogenase: toward a mechanism of enzymatic nitrogen fixation, *Acc Chem Res 42*, 609-619.

31. Igarashi, R. Y., and Seefeldt, L. C. (2003) Nitrogen fixation: the mechanism of the Mo-dependent nitrogenase, *Crit Rev Biochem Mol Biol 38*, 351-384.

32. Peters, J. W., Fisher, K., and Dean, D. R. (1995) Nitrogenase structure and function: a biochemical-genetic perspective, *Annu. Rev. Microbiol. 49*, 335-366.

33. Peters, J. W., and Szilagyi, R. K. (2006) Exploring new frontiers of nitrogenase structure and mechanism, *Curr Opin Chem Biol 10*, 101-108.

34. Rees, D. C., and Howard, J. B. (2000) Nitrogenase: standing at the crossroads, *Curr Opin Chem Biol 4*, 559-566.

35. Seefeldt, L. C., Hoffman, B. M., and Dean, D. R. (2009) Mechanism of Mo-dependent nitrogenase, *Annu Rev Biochem 78*, 701-722.

36. Lee, S. C., and Holm, R. H. (2004) The clusters of nitrogenase: synthetic methodology in the construction of weak-field clusters, *Chem Rev 104*, 1135-1158.

37. Eady, R. R., Robson, R. L., Richardson, T. H., Miller, R. W., and Hawkins, M. (1987) The vanadium nitrogenase of Azotobacter chroococcum. Purification and properties of the VFe protein, *The Biochemical journal 244*, 197-207.

38. Eady, R. R., Smith, B. E., Cook, K. A., and Postgate, J. R. (1972) Nitrogenase of Klebsiella pneumoniae. Purification and properties of the component proteins, *The Biochemical journal 128*, 655-675.

39. Schneider, K., Gollan, U., Selsemeier-Voigt, S., Plass, W., and Muller, A. (1994) Rapid purification of the protein components of a highly active "iron only" nitrogenase, *Die Naturwissenschaften 81*, 405-408.

40. Schneider, K., Gollan, U., Drottboom, M., Selsemeier-Voigt, S., and Muller, A. (1997) Comparative biochemical characterization of the iron-only nitrogenase and the molybdenum nitrogenase from Rhodobacter capsulatus, *European journal of biochemistry / FEBS 244*, 789-800.

41. Poole, R. K., and Hill, S. (1997) Respiratory protection of nitrogenase activity in Azotobacter vinelandii--roles of the terminal oxidases, *Biosci Rep 17*, 303-317.

42. Maier, R. J., and Moshiri, F. (2000) Role of the Azotobacter vinelandii nitrogenase-protective shethna protein in preventing oxygen-mediated cell death, *J Bacteriol 182*, 3854-3857.

43. Ureta, A., and Nordlund, S. (2002) Evidence for conformational protection of nitrogenase against oxygen in Gluconacetobacter diazotrophicus by a putative FeSII protein, *J Bacteriol 184*, 5805-5809.

44. Vance, C. P. (2001) Symbiotic nitrogen fixation and phosphorus acquisition. Plant nutrition in a world of declining renewable resources, *Plant physiology 127*, 390-397.

45. Downie, J. A. (2005) Legume haemoglobins: symbiotic nitrogen fixation needs bloody nodules, *Current biology : CB 15*, R196-198.

46. Fay, P. (1992) Oxygen relations of nitrogen fixation in cyanobacteria, *Microbiol Rev 56*, 340-373.

47. Dixon, R., Eady, R. R., Espin, G., Hill, S., Iaccarino, M., Kahn, D., and Merrick, M. (1980) Analysis of regulation of Klebsiella pneumoniae nitrogen fixation (nif) gene cluster with gene fusions, *Nature 286*, 128-132.

48. Stal, L. J., and Krumbien, W. E. (1985) Nitrogenase activity in the non-heterocystous cyanobacterium Oscillatoria sp. grown under alternating light-dark cycles, *Arch Microbiol 143*, 67-71.

49. Vandecasteele, J. P., and Burris, R. H. (1970) Purification and properties of the constituents of the nitrogenase complex from Clostridium pasteurianum, *J Bacteriol 101*, 794-801.

50. Georgiadis, M. M., Komiya, H., Chakrabarti, P., Woo, D., Kornuc, J. J., and Rees, D. C. (1992) Crystallographic structure of the nitrogenase iron protein from *Azotobacter vinelandii*, *Science 257*, 1653-1659.

51. Orme-Johnson, W. H., Hamilton, W. D., Jones, T. L., Tso, M. Y., Burris, R. H., Shah, V. K., and Brill, W. J. (1972) Electron paramagnetic resonance of nitrogenase and nitrogenase components from Clostridium pasteurianum W5 and Azotobacter vinelandii OP, *Proc Natl Acad Sci U S A 69*, 3142-3145.

52. Emerich, D. W., Hageman, R. V., and Burris, R. H. (1981) Interactions of dinitrogenase and dinitrogenase reductase, *Advances in enzymology and related areas of molecular biology 52*, 1-22.

53. Thorneley, R. N., and Lowe, D. J. (1983) Nitrogenase of Klebsiella pneumoniae. Kinetics of the dissociation of oxidized iron protein from molybdenum-iron protein: identification of the rate-limiting step for substrate reduction, *The Biochemical journal 215*, 393-403.

54. Thorneley, R. N., and Lowe, D. J. (1984) The mechanism of Klebsiella pneumoniae nitrogenase action. Simulation of the dependences of H2-evolution rate on component-protein concentration and ratio and sodium dithionite concentration, *The Biochemical journal 224*, 903-909.

55. Lanzilotta, W. N., and Seefeldt, L. C. (1997) Changes in the midpoint potentials of the nitrogenase metal centers as a result of iron protein-molybdenum-iron protein complex formation, *Biochemistry 36*, 12976-12983.

56. Koonin, E. V. (1993) A superfamily of ATPases with diverse functions containing either classical or deviant ATP-binding motif, *J Mol Biol 229*, 1165-1174.

57. Schulze, G. E. (1992) Binding of nucleotides by proteins: Current opinion in structural biology *Current Opinion in Structural Biology 2*, 61 - 67.

58. Schlessman, J. L., Woo, D., Joshua-Tor, L., Howard, J. B., and Rees, D. C. (1998) Conformational variability in structures of the nitrogenase iron proteins from Azotobacter vinelandii and Clostridium pasteurianum, *J Mol Biol 280*, 669-685.

59. Locher, K. P., Hans, M., Yeh, A. P., Schmid, B., Buckel, W., and Rees, D. C. (2001) Crystal structure of the Acidaminococcus fermentans 2-hydroxyglutaryl-CoA dehydratase component A, *J Mol Biol 307*, 297-308.

60. Sarma, R., Barney, B. M., Hamilton, T. L., Jones, A., Seefeldt, L. C., and Peters, J. W. (2008) Crystal structure of the L protein of Rhodobacter sphaeroides light-independent protochlorophyllide reductase with MgADP bound: a homologue of the nitrogenase Fe protein, *Biochemistry 47*, 13004-13015.

61. Burke, D. H., Hearst, J. E., and Sidow, A. (1993) Early evolution of photosynthesis: clues from nitrogenase and chlorophyll iron proteins, *Proc Natl Acad Sci U S A 90*, 7134-7138.

62. Howard, J. B., and Rees, D. C. (1994) Nitrogenase: a nucleotide-dependent molecular switch, *Annu Rev Biochem 63*, 235-264.

63. Seefeldt, L. C., and Dean, D. R. (1997) Role of Nucleotides in Nitrogenase Catalysis, *Accounts of Chemical Research 30*, 260 - 266.

64. Chiu, H., Peters, J. W., Lanzilotta, W. N., Ryle, M. J., Seefeldt, L. C., Howard, J. B., and Rees, D. C. (2001) MgATP-Bound and nucleotide-free structures of a nitrogenase protein complex between the Leu 127 Delta-Fe-protein and the MoFe-protein, *Biochemistry 40*, 641-650.

65. Schindelin, H., Kisker, C., Schlessman, J. L., Howard, J. B., and Rees, D. C. (1997) Structure of ADP x AIF4(-)-stabilized nitrogenase complex and its implications for signal transduction, *Nature 387*, 370-376.

66. Tezcan, F. A., Kaiser, J. T., Mustafi, D., Walton, M. Y., Howard, J. B., and Rees, D. C. (2005) Nitrogenase complexes: multiple docking sites for a nucleotide switch protein, *Science 309*, 1377-1380.

67. Chan, J. M., Wu, W., Dean, D. R., and Seefeldt, L. C. (2000) Construction and characterization of a heterodimeric iron protein: defining roles for adenosine triphosphate in nitrogenase catalysis, *Biochemistry 39*, 7221-7228.

68. Lanzilotta, W. N., Ryle, M. J., and Seefeldt, L. C. (1995) Nucleotide hydrolysis and protein conformational changes in Azotobacter vinelandii nitrogenase iron protein: defining the function of aspartate 129, *Biochemistry 34*, 10713-10723.

69. Lindahl, P. A., Day, E. P., Kent, T. A., Orme-Johnson, W. H., and Munck, E. (1985) Mossbauer, EPR, and magnetization studies of the Azotobacter vinelandii Fe protein. Evidence for a [4Fe-4S]1+ cluster with spin S = 3/2, *J Biol Chem 260*, 11160-11173.

70. Lindahl, P. A., Gorelick, N. J., Munck, E., and Orme-Johnson, W. H. (1987) EPR and Mossbauer studies of nucleotide-bound nitrogenase iron protein from Azotobacter vinelandii, *J Biol Chem 262*, 14945-14953.

71. Morgan, T. V., Prince, R. C., and Mortenson, L. E. (1986) Electrochemical titration of the S=3/2 and S=1/2 states of the iron of nitrogenase, *FEBS LETTERS 206*, 4-8.

72. Ryle, M. J., Lanzilotta, W. N., Mortenson, L. E., Watt, G. D., and Seefeldt, L. C. (1995) Evidence for a central role of lysine 15 of Azotobacter vinelandii

nitrogenase iron protein in nucleotide binding and protein conformational changes, *J Biol Chem 270*, 13112-13117.

73. Ryle, M. J., Lanzilotta, W. N., and Seefeldt, L. C. (1996) Elucidating the mechanism of nucleotide-dependent changes in the redox potential of the [4Fe-4S] cluster in nitrogenase iron protein: the role of phenylalanine 135, *Biochemistry 35*, 9424-9434.

74. Ryle, M. J., Lanzilotta, W. N., Seefeldt, L. C., Scarrow, R. C., and Jensen, G. M. (1996) Circular dichroism and x-ray spectroscopies of Azotobacter vinelandii nitrogenase iron protein. MgATP and MgADP induced protein conformational changes affecting the [4Fe-4S] cluster and characterization of a [2Fe-2S] form, *J Biol Chem 271*, 1551-1557.

75. Stephens, P. J., McKenna, C. E., Smith, B. E., Nguyen, H. T., McKenna, M. C., Thomson, A. J., Devlin, F., and Jones, J. B. (1979) Circular dichroism and magnetic circular dichroism of nitrogenase proteins, *Proc Natl Acad Sci U S A 76*, 2585-2589.

76. Fani, R., Gallo, R., and Lio, P. (2000) Molecular evolution of nitrogen fixation: the evolutionary history of the nifD, nifK, nifE, and nifN genes, *J Mol Evol 51*, 1-11.

77. Bolin, J. T., Ronco, A. E., Morgan, T. V., Mortenson, L. E., and Xuong, N. H. (1993) The unusual metal clusters of nitrogenase: structural features revealed by x-ray anomalous diffraction studies of the MoFe protein from Clostridium pasteurianum, *Proc Natl Acad Sci U S A 90*, 1078-1082.

78. Chan, M. K., Kim, J., and Rees, D. C. (1993) The nitrogenase FeMo-cofactor and P-cluster pair: 2.2 A resolution structures, *Science 260*, 792-794.

79. Einsle, O., Tezcan, F. A., Andrade, S. L., Schmid, B., Yoshida, M., Howard, J. B., and Rees, D. C. (2002) Nitrogenase MoFe-protein at 1.16 A resolution: a central ligand in the FeMo-cofactor, *Science 297*, 1696-1700.

80. Kim, J., and Rees, D. C. (1992) Crystallographic structure and functional implications of the nitrogenase molybdenum-iron protein from *Azotobacter vinelandii*, *Nature 360*, 553-560.

81. Kim, J., Woo, D., and Rees, D. C. (1993) X-ray crystal structure of the nitrogenase molybdenum-iron protein from Clostridium pasteurianum at 3.0-A resolution, *Biochemistry 32*, 7104-7115.

82. Mayer, S. M., Lawson, D. M., Gormal, C. A., Roe, S. M., and Smith, B. E. (1999) New insights into structure-function relationships in nitrogenase: A 1.6 A

resolution X-ray crystallographic study of Klebsiella pneumoniae MoFe-protein, *J Mol Biol 292*, 871-891.

83. Peters, J. W., Stowell, M. H., Soltis, S. M., Finnegan, M. G., Johnson, M. K., and Rees, D. C. (1997) Redox-dependent structural changes in the nitrogenase P-cluster, *Biochemistry 36*, 1181-1187.

84. Brocker, M. J., Schomburg, S., Heinz, D. W., Jahn, D., Schubert, W. D., and Moser, J. (2010) Crystal structure of the nitrogenase-like dark operative protochlorophyllide oxidoreductase catalytic complex (ChlN/ChlB)2, *J Biol Chem 285*, 27336-27345.

85. Brocker, M. J., Virus, S., Ganskow, S., Heathcote, P., Heinz, D. W., Schubert, W. D., Jahn, D., and Moser, J. (2008) ATP-driven reduction by dark-operative protochlorophyllide oxidoreductase from Chlorobium tepidum mechanistically resembles nitrogenase catalysis, *J Biol Chem 283*, 10559-10567.

86. Brocker, M. J., Watzlich, D., Uliczka, F., Virus, S., Saggu, M., Lendzian, F., Scheer, H., Rudiger, W., Moser, J., and Jahn, D. (2008) Substrate recognition of nitrogenase-like dark operative protochlorophyllide oxidoreductase from Prochlorococcus marinus, *J Biol Chem 283*, 29873-29881.

87. Muraki, N., Nomata, J., Ebata, K., Mizoguchi, T., Shiba, T., Tamiaki, H., Kurisu, G., and Fujita, Y. (2010) X-ray crystal structure of the light-independent protochlorophyllide reductase, *Nature 465*, 110-114.

88. Fujita, Y. (1996) Protochlorophyllide reduction: a key step in the greening of plants, *Plant & cell physiology 37*, 411-421.

89. Fujita, Y., and Bauer, C. E. (2000) Reconstitution of light-independent protochlorophyllide reductase from purified bchl and BchN-BchB subunits. In vitro confirmation of nitrogenase-like features of a bacteriochlorophyll biosynthesis enzyme, *J Biol Chem 275*, 23583-23588.

90. Fujita, Y., Takagi, H., and Hase, T. (1998) Cloning of the gene encoding a protochlorophyllide reductase: the physiological significance of the co-existence of light-dependent and -independent protochlorophyllide reduction systems in the cyanobacterium Plectonema boryanum, *Plant & cell physiology 39*, 177-185.

91. Einsle, O., Tezcan, F. A., Andrade, S. L. A., Schmid, B., Yoshida, M., Howard, J. B., and Rees, D. C. (2002) Nitrogenase MoFe-protein at 1.16 A resolution: a central ligand in the FeMo-cofactor, *Science 297*, 1696-1700.

92. Kim, J., and Rees, D. C. (1992) Structural models for the metal centers in the nitrogenase molybdenum-iron protein, *Science 257*, 1677-1682.

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

94. Lowe, D. J., Fisher, K., and Thorneley, R. N. (1993) Klebsiella pneumoniae nitrogenase: pre-steady-state absorbance changes show that redox changes occur in the MoFe protein that depend on substrate and component protein ratio; a role for P-centres in reducing dinitrogen?, *The Biochemical journal 292 (Pt 1)*, 93-98.

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

96. Kent, H. M., Ioannidis, I., Gormal, C., Smith, B. E., and Buck, M. (1989) Site-directed mutagenesis of the Klebsiella pneumoniae nitrogenase. Effects of modifying conserved cysteine residues in the alpha- and beta-subunits, *The Biochemical journal 264*, 257-264.

97. Kent, H. M., Baines, M., Gormal, C., Smith, B. E., and Buck, M. (1990) Analysis of site-directed mutations in the alpha- and beta-subunits of Klebsiella pneumoniae nitrogenase, *Molecular microbiology 4*, 1497-1504.

98. Dean, D. R., Setterquist, R. A., Brigle, K. E., Scott, D. J., Laird, N. F., and Newton, W. E. (1990) Evidence that conserved residues Cys-62 and Cys-154 within the Azotobacter vinelandii nitrogenase MoFe protein alpha-subunit are essential for nitrogenase activity but conserved residues His-83 and Cys-88 are not, *Molecular microbiology 4*, 1505-1512.

99. Scott, D. J., May, H. D., Newton, W. E., Brigle, K. E., and Dean, D. R. (1990) Role for the nitrogenase MoFe protein alpha-subunit in FeMo-cofactor binding and catalysis, *Nature 343*, 188-190.

100. Scott, D. J., Dean, D. R., and Newton, W. E. (1992) Nitrogenase-catalyzed ethane production and CO-sensitive hydrogen evolution from MoFe proteins having amino acid substitutions in an alpha-subunit FeMo cofactor-binding domain, *J Biol Chem 267*, 20002-20010.

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

102. Jacobson, M. R., Brigle, K. E., Bennett, L. T., Setterquist, R. A., Wilson, M. S., Cash, V. L., Beynon, J., Newton, W. E., and Dean, D. R. (1989) Physical and genetic map of the major nif gene cluster from Azotobacter vinelandii, *J Bacteriol 171*, 1017-1027.

103. Jacobson, M. R., Cash, V. L., Weiss, M. C., Laird, N. F., Newton, W. E., and Dean, D. R. (1989) Biochemical and genetic analysis of the nifUSVWZM cluster from Azotobacter vinelandii, *Mol Gen Genet 219*, 49-57.

104. Joerger, R. D., and Bishop, P. E. (1988) Nucleotide sequence and genetic analysis of the nifB-nifQ region from Azotobacter vinelandii, *J Bacteriol 170*, 1475-1487.

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

106. Curatti, L., Brown, C. S., Ludden, P. W., and Rubio, L. M. (2005) Genes required for rapid expression of nitrogenase activity in Azotobacter vinelandii, *Proc Natl Acad Sci U S A 102*, 6291-6296.

107. Ugalde, R. A., Imperial, J., Shah, V. K., and Brill, W. J. (1984) Biosynthesis of iron-molybdenum cofactor in the absence of nitrogenase, *J Bacteriol 159*, 888-893.

108. White, T. C., Harris, G. S., and Orme-Johnson, W. H. (1992) Electrophoretic studies on the assembly of the nitrogenase molybdenum-iron protein from the Klebsiella pneumoniae nifD and nifK gene products, *J Biol Chem 267*, 24007-24016.

109. Johnson, D. C., Dos Santos, P. C., and Dean, D. R. (2005) NifU and NifS are required for the maturation of nitrogenase and cannot replace the function of isc-gene products in Azotobacter vinelandii, *Biochemical Society transactions 33*, 90-93.

110. Smith, A. D., Jameson, G. N., Dos Santos, P. C., Agar, J. N., Naik, S., Krebs, C., Frazzon, J., Dean, D. R., Huynh, B. H., and Johnson, M. K. (2005) NifS-mediated assembly of [4Fe-4S] clusters in the N- and C-terminal domains of the NifU scaffold protein, *Biochemistry 44*, 12955-12969.

111. Zheng, L., and Dean, D. R. (1994) Catalytic formation of a nitrogenase iron-sulfur cluster, *J Biol Chem 269*, 18723-18726.

112. Dos Santos, P. C., Smith, A. D., Frazzon, J., Cash, V. L., Johnson, M. K., and Dean, D. R. (2004) Iron-sulfur cluster assembly: NifU-directed activation of the nitrogenase Fe protein, *J Biol Chem 279*, 19705-19711.

113. Corbett, M. C., Hu, Y., Fay, A. W., Tsuruta, H., Ribbe, M. W., Hodgson, K. O., and Hedman, B. (2007) Conformational differences between Azotobacter vinelandii nitrogenase MoFe proteins as studied by small-angle X-ray scattering, *Biochemistry 46*, 8066-8074.

114. Magnuson, J. K., Paustian, T. D., Shah, V. K., Dean, D. R., Roberts, G. P., Rees, D. C., and Howard, J. B. (1997) Nitrogenase iron-molybdenum cofactor binding site: Protein conformational changes associated with cofactor binding, *Tetrahedron 53*, 11971-11984.

115. Allen, R. M., Homer, M. J., Chatterjee, R., Ludden, P. W., Roberts, G. P., and Shah, V. K. (1993) Dinitrogenase reductase- and MgATP-dependent maturation of apodinitrogenase from Azotobacter vinelandii, *J Biol Chem 268*, 23670-23674.

116. Hu, Y., Fay, A. W., Lee, C. C., and Ribbe, M. W. (2007) P-cluster maturation on nitrogenase MoFe protein, *Proc Natl Acad Sci U S A 104*, 10424-10429.

117. Rubio, L. M., Rangaraj, P., Homer, M. J., Roberts, G. P., and Ludden, P. W. (2002) Cloning and mutational analysis of the gamma gene from Azotobacter vinelandii defines a new family of proteins capable of metallocluster binding and protein stabilization, *J Biol Chem 277*, 14299-14305.

118. Boyd, E. S., Anbar, A. D., Miller, S., Hamilton, T. L., Lavin, M., and Peters, J. W. (2011) A Late Methanogen Origin for Molybdenum-Dependent Nitrogenase, *Geiobiol*.

119. Kaiser, J. T., Hu, Y., Wiig, J. A., Rees, D. C., and Ribbe, M. W. (2011) Structure of precursor-bound NifEN: a nitrogenase FeMo cofactor maturase/insertase, *Science 331*, 91-94.

120. Sofia, H. J., Chen, G., Hetzler, B. G., Reyes-Spindola, J. F., and Miller, N. E. (2001) Radical SAM, a novel protein superfamily linking unresolved steps in familiar biosynthetic pathways with radical mechanisms: functional characterization using new analysis and information visualization methods, *Nucleic acids research 29*, 1097-1106.

121. Curatti, L., Ludden, P. W., and Rubio, L. M. (2006) NifB-dependent in vitro synthesis of the iron-molybdenum cofactor of nitrogenase, *Proc Natl Acad Sci U S A 103*, 5297-5301.

122. Wiig, J. A., Hu, Y., and Ribbe, M. W. (2011) NifEN-B complex of Azotobacter vinelandii is fully functional in nitrogenase FeMo cofactor assembly, *Proc Natl Acad Sci U S A 108*, 8623-8627.

123. Hernandez, J. A., Curatti, L., Aznar, C. P., Perova, Z., Britt, R. D., and Rubio, L. M. (2008) Metal trafficking for nitrogen fixation: NifQ donates molybdenum to NifEN/NifH for the biosynthesis of the nitrogenase FeMo-cofactor, *Proc Natl Acad Sci U S A 105*, 11679-11684.

124. Rangaraj, P., and Ludden, P. W. (2002) Accumulation of 99Mo-containing iron-molybdenum cofactor precursors of nitrogenase on NifNE, NifH, and NifX of Azotobacter vinelandii, *J Biol Chem 277*, 40106-40111.

125. Brill, W. J. (1980) Biochemical genetics of nitrogen fixation, *Microbiol Rev 44*, 449-467.

126. Brigle, K. E., Newton, W. E., and Dean, D. R. (1985) Complete nucleotide sequence of the Azotobacter vinelandii nitrogenase structural gene cluster, *Gene 37*, 37-44.

127. Kessler, P. S., Blank, C., and Leigh, J. A. (1998) The nif gene operon of the methanogenic archaeon Methanococcus maripaludis, *J Bacteriol 180*, 1504-1511.

128. Bishop, P. E., Jarlenski, D. M., and Hetherington, D. R. (1980) Evidence for an alternative nitrogen fixation system in Azotobacter vinelandii, *Proc Natl Acad Sci U S A 77*, 7342-7346.

129. Martinez-Argudo, I., Little, R., Shearer, N., Johnson, P., and Dixon, R. (2005) Nitrogen fixation: key genetic regulatory mechanisms, *Biochemical Society transactions 33*, 152-156.

130. Belay, N., Sparling, R., and Daniels, L. (1984) Dinitrogen fixation by a thermophilic methanogenic bacterium, *Nature 312*, 286-288.

131. Steunou, A. S., Bhaya, D., Bateson, M. M., Melendrez, M. C., Ward, D. M., Brecht, E., Peters, J. W., Kuhl, M., and Grossman, A. R. (2006) In situ analysis of nitrogen fixation and metabolic switching in unicellular thermophilic cyanobacteria inhabiting hot spring microbial mats, *Proc Natl Acad Sci U S A 103*, 2398-2403.

132. Bekker, A., Holland, H. D., Wang, P. L., Rumble, D., 3rd, Stein, H. J., Hannah, J. L., Coetzee, L. L., and Beukes, N. J. (2004) Dating the rise of atmospheric oxygen, *Nature 427*, 117-120.

133. Kasting, J. F. (2001) Earth history. The rise of atmospheric oxygen, *Science 293*, 819-820.

134. Pilcher, C. B. (2003) Biosignatures of early earths, *Astrobiology 3*, 471-486.

135. Javaux, E. J. (2006) Extreme life on Earth--past, present and possibly beyond, *Research in microbiology 157*, 37-48.

136. Wachtershauser, G. (2006) From volcanic origins of chemoautotrophic life to Bacteria, Archaea and Eukarya, *Philosophical transactions of the Royal Society of London. Series B, Biological sciences 361*, 1787-1806; discussion 1806-1788.

137. Gaucher, E. A., Govindarajan, S., and Ganesh, O. K. (2008) Palaeotemperature trend for Precambrian life inferred from resurrected proteins, *Nature 451*, 704-707.

138. Shen, Y., Buick, R., and Canfield, D. E. (2001) Isotopic evidence for microbial sulphate reduction in the early Archaean era, *Nature 410*, 77-81.

139. Reysenbach, A. L., and Shock, E. (2002) Merging genomes with geochemistry in hydrothermal ecosystems, *Science 296*, 1077-1082.

140. Michaelis, W., Seifert, R., Nauhaus, K., Treude, T., Thiel, V., Blumenberg, M., Knittel, K., Gieseke, A., Peterknecht, K., Pape, T., Boetius, A., Amann, R., Jorgensen, B. B., Widdel, F., Peckmann, J., Pimenov, N. V., and Gulin, M. B. (2002) Microbial reefs in the Black Sea fueled by anaerobic oxidation of methane, *Science 297*, 1013-1015.

141. Shirey, S. B., and Richardson, S. H. (2011) Start of the Wilson Cycle at 3 Ga Shown by Diamonds from Subcontinental Mantle, *Science 333*, 434-436.

142. Gribaldo, S., and Brochier-Armanet, C. (2006) The origin and evolution of Archaea: a state of the art, *Philosophical transactions of the Royal Society of London. Series B, Biological sciences 361*, 1007-1022.

143. Reysenbach, A. L., and Cady, S. L. (2001) Microbiology of ancient and modern hydrothermal systems, *Trends in microbiology 9*, 79-86.

144. Giulio, M. (2003) The Universal Ancestor and the Ancestor of Bacteria Were Hyperthermophiles, *Journal of Molecular Evolution 57*, 721 - 730.

145. Walter, M. R., and Des Marais, D. J. (1993) Preservation of biological information in thermal spring deposits: developing a strategy for the search for fossil life on Mars, *Icarus 101*, 129-143.

146. Amend, J. P., and Shock, E. L. (2001) Energetics of overall metabolic reactions of thermophilic and hyperthermophilic Archaea and bacteria, *FEMS microbiology reviews 25*, 175-243.

147. Boyd, E. S., Leavitt, W. D., and Geesey, G. G. (2009) CO(2) uptake and fixation by a thermoacidophilic microbial community attached to precipitated sulfur in a geothermal spring, *Appl Environ Microbiol 75*, 4289-4296.

148. Spear, J. R., Walker, J. J., McCollom, T. M., and Pace, N. R. (2005) Hydrogen and bioenergetics in the Yellowstone geothermal ecosystem, *Proc Natl Acad Sci U S A 102*, 2555-2560.

149. Jackson, C. R., Langner, H. W., Donahoe-Christiansen, J., Inskeep, W. P., and McDermott, T. R. (2001) Molecular analysis of microbial community structure in an arsenite-oxidizing acidic thermal spring, *Environmental microbiology 3*, 532-542.

150. Hamamura, N., Olson, S. H., Ward, D. M., and Inskeep, W. P. (2005) Diversity and functional analysis of bacterial communities associated with natural hydrocarbon seeps in acidic soils at Rainbow Springs, Yellowstone National Park, *Appl Environ Microbiol 71*, 5943-5950.

151. Ward, D. M., Ferris, M. J., Nold, S. C., and Bateson, M. M. (1998) A natural view of microbial biodiversity within hot spring cyanobacterial mat communities, *Microbiology and molecular biology reviews : MMBR 62*, 1353-1370.

152. Barns, S. M., Delwiche, C. F., Palmer, J. D., and Pace, N. R. (1996) Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences, *Proc Natl Acad Sci U S A 93*, 9188-9193.

153. Huber, H., Hohn, M. J., Rachel, R., Fuchs, T., Wimmer, V. C., and Stetter, K. O. (2002) A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont, *Nature 417*, 63-67.

CHAPTER 2

EXAMINING THE ORIGIN AND EVOLUTION OF BIOLOGICAL NITROGEN FIXATION

Manuscripts in Chapter 2

Contributions of Authors and Co-Authors

Author: Eric S. Boyd

Contributions: Conceived the study, collected and analyzed output data, and wrote the manuscript.

Co-author: Trinity L. Hamilton

Contributions: Collected output data, discussed results and implications and edited the manuscript.

Co-author: Ariel D. Anbar

Contributions: Discussed results and implications and edited the manuscript.

Co-author: Scott Miller

Contributions: Discussed the results and implications, analyzed output data, and edited the manuscript.

Co-author: Matthew Lavin

Contributions: Discussed the results and implications analyzed output data, and edited the manuscript.

Co-author: John W. Peters

Contributions: Obtained funding, discussed the results and implications and edited the manuscript.

Manuscript Information Page

A LATE METHANOGEN ORIGIN FOR MOLYBDENUM-DEPENDENT NITROGENASE

Authors: Eric S. Boyd, Ariel D. Anbar, Scott Miller, Trinity L. Hamilton, Matthew Lavin, and John W. Peters

Journal: Geobiology

Status of the Manuscript:

___Prepared for submission to a peer-reviewed journal
___Officially submitted to a peer-reviewed journal
___Accepted by a peer-reviewed journal
_x_Published in a peer-reviewed journal

Publisher: Wiley Online Library

Issue in which the manuscript appears: Volume 9, 221 – 232 (2011)

AN ALTERNATIVE PATH FOR THE EVOLUTION OF BIOLOGICAL NITROGEN FIXATION

Authors: Eric S. Boyd, Trinity L. Hamilton, John W. Peters

Journal: Frontiers in Microbiology

Status of the Manuscript:

___Prepared for submission to a peer-reviewed journal
___Officially submitted to a peer-reviewed journal
___Accepted by a peer-reviewed journal
_x_Published in a peer-reviewed journal

Publisher: Frontiers Community

First published October 2011, doi: 10.3389/fmicb.2011.00205

CHAPTER 2

EXAMINING THE ORIGIN AND EVOLUTION OF BIOLOGICAL NITROGEN FIXATION

Introduction

Biologically available nitrogen (fixed N) is essential to all life. Initially, abiotic fixed N such as from lightening discharge were the only sources for early life (*1*, *2*). As the biome expanded on early Earth, however, abiotic sources of fixed N presumably became limiting, leading to the innovation of biological nitrogen fixation, the conversion of atmospheric dinitrogen (N_2) to biologically available ammonia. Nitrogenase, the enzyme that catalyzes dinitrogen reduction, would have had a profound impact on the evolution and proliferation of life and the biogeochemical cycles that it modulates, as biological nitrogen fixation would relieve fixed N limitation (*3*, *4*). Prior to the rise of oxygen in the atmosphere ~2.5 Ga, it is thought that the biological nitrogen fixation would have been catalyzed by an alternative form of nitrogenase that did not use Mo, because the oceans were depleted in Mo due to the insolubility of Mo-sulfides (*5*-*7*). Therefore, the ancestral forms of the nitrogenase enzyme were thought to contain V and Fe (Vnf) or only Fe (Anf) at the active site metal cofactor; however, it has remained unclear which form of the enzyme was the first to emerge (*8*, *9*).

The predominance of the Mo-nitrogenase in diazotrophs present today and the conserved nature of the enzyme components has enabled in-depth phylogenetic examination of the deduced amino acid sequences of nitrogenase. Recently, these studies

lead to two equally parsimonious origins for the Mo-containing form of the enzyme. The first, the LUCA origin model, proposes that Mo-nitrogenase was present in the last universal common ancestor (LUCA) (*10*) while the second, the methanogen origin model, invokes an origin for Mo-nitrogenase in the methanogenic archaea with the bacterial domain acquiring the enzyme as a result of lateral gene transfer (LGT) (*11*). Both proposals require that Nif emerged in an anoxic environment. The LUCA model invokes an origin of Mo-nitrogenase in anoxic oceans prior to the rise of O_2 because O_2 was likely scarce until roughly ~2.5 Ga (*5, 12, 13*), which postdates the LUCA and the subsequent split to form bacteria and archaea (*14-16*). All methanogens are obligate anaerobes (*17, 18*), therefore the methanogen origin model suggests the emergence of the enzyme in methanogens would have occurred under strictly anaerobic conditions. The origin of Mo-nitrogenase in an anoxic environment proposed by both scenarios is consistent with the extreme oxygen sensitivity of the nitrogenase enzyme.

The conserved nature of both the mechanism of nitrogen fixation and the sequence homology of component proteins makes it possible to explore the evolution and origin of both processes. Unfortunately, the frequency of horizontal transfer of the nitrogenase enzyme operon makes the phylogenetic analysis quite difficult. In addition, the presence of the alternative nitrogenases which are related but differ in active site ligand composition, make inferring evolutionary relationships more difficult. The recent increase in publically available genomic sequences has aided in deducing the evolutionary history of nitrogen fixation and photosynthesis. Genomic sequences have facilitated the compilation of a large number of nitrogenase genes and uncharacterized

sequences which exhibit homology to these sequences. Phylogenetic reconstructions have resolved these sequences into five distinct clades. These five clades are aerobic Mo-dependent nitrogenases, anaerobic Mo-nitrogenases, alternative nitrogenases, uncharacterized *nif* homologs, and (bacterio)chlorophyll biosynthesis genes. The typical Mo-dependent nitrogenase clade is populated by mostly cyanobacteria and proteobacteria, while methanogens, clostridia, and sulfate-reducing bacteria are the predominant taxa of the anaerobic clade (*11*). Some hypotheses claim the alternative nitrogenases emerged first and selective pressure for the more efficient Mo-nitrogenase became prevalent as molybdenum became more biologically available as a result of the oxygenation of the Earth.

The uncharacterized *nif* homologs, *nfl's* (nif-like), are present in some methanogens, including some that don't fix nitrogen, and some diazotrophic bacteria capable of anoxygenic photosynthesis (*11*). There are two *nfl* homologs, *nflH* and *nflD*, that are homologous to *nifH* and *nifD.* The evolutionary history of these homologs also remains unknown, as does their function although they have been hypothesized to be involved in F430 biosynthesis, a nickel porphinoid in methyl coenzyme reductase, a key enzyme in methanogenesis (*19*). Preliminary evidence reveals that the expression of the *nflH* and *nflD* gene products are not regulated by nitrogen and are not involved in nitrogen fixation. The NflH protein is related to NifH and the BchL proteins and is especially conserved in nucleotide-binding regions and the conserved cysteine residues responsible for metal cluster binding. This sequence conservation, as well as the location of *nfl* sequences basal to *nif* and (b)chl sequences in phylogenetic analyses, suggests a gene duplication of the *nflH* and *nflD* may have given rise to the genetic machinery

necessary for nitrogen fixation and bacteriochlorophyll biosynthesis, or that these two may have diverged from the *nfls*.

The majority of present-day biological nitrogen fixation is catalyzed by the Mo-dependent nitrogenase, encoded by *nif (20)*. However, the alternative forms of the enzyme, V-nitrogenase, encoded by *vnf*, and Fe-nitrogenase, encoded by *anf*, can serve as important sources of fixed nitrogen in environments were Mo is limiting (*21*). To date, the alternative nitrogenases have only been found in the genomes of diazotrophs that also encode a Mo-nitrogenase (*11, 22*). The Mo-nitrogenase consists the Fe protein (NifH), a homodimer bridged by a [4Fe-4S] cluster, that serves as the obligate electron donor to the MoFe protein (NifDK). NifDK is a heterotetramer which houses the complex metal cofactor (FeMo-co) that is the site of substrate reduction. *nif* regulons encode the structural proteins as well as a suite of genes involved in regulation and biosynthesis of FeMo-co (*20*). Several of these genes are thought to have evolved from duplication and fusion events, namely *nifE* and *nifN (11, 22, 23)*, which encode the scaffold protein, NifEN where FeMo-co is assembled, as well as *nifB* which encodes a radical S-adenosyl methionine (SAM) enzyme (NifB) that catalyzes the formation of NifB-co, a metal cluster precursor to FeMo-co (*20, 24*).

Geochemical and isotopic evidence provides clues for the emergence of biological nitrogen fixation in the evolution and proliferation of life on Earth. From this data, it has been proposed that the alternative forms of the enzyme were the first to emerge because ancient oceans were limited in soluble Mo but Fe-rich prior to the rise of O_2. Phylogenetic evidence has lead to two proposals for the emergence of the Mo-nitrogenase, both of which invoke the enzyme emerging under anoxic conditions with the

LUCA model hypothesizing the presence of Mo-nitrogenase prior to the split of bacteria and archaea and the methanogen origin model proposing the emergence of Mo-nitrogenase in methanogens followed by subsequent lateral transfer to bacteria.

The combination of geochemical evidence with the increasing amount of available genomic sequence data has enabled a better understanding of the evolution and emergence of this key nitrogen cycling enzyme. Phylogenetic analyses examining the concatenations of protein homologs of the structural components (H, D and K) shared between all known nitrogenases have led to an alternative hypothesis for the trajectory of specific metal incorporation into the active site cofactor during the evolution of nitrogenase. This phylogenetic- and structure-based analysis indicates an evolutionary path whereby the Mo-nitrogenase emerged within the methanogenic archaea and then gave rise to the alternative forms of the enzymes, perhaps with local Mo limitation providing the selective pressure to do so. The structural-based examination of nitrogenase and nitrogenase homologs suggests that the ancestor of all nitrogenases had an open cavity capable of binding metal clusters which conferred reactivity. The availability of fixed nitrogen in combination with local environmental conditions affecting metal availability would have presumably controlled evolution of this ancestral form of the enzyme. When Mo became sufficiently bioavailable evolution of the enzyme would have refined the cofactor and enzyme to the more common form (Mo-nitrogenase) that is perpetuated in extant biology. To further examine the evolution of the Mo-nitrogenase, a phylogenetic approach was used to examine the gene duplication and gene fusion events that gave rise to genes necessary for synthesis of FeMo-co. These analyses examined the presumed paralogous gene duplication of *nifDK* to *nifEN* and the gene fusion event of

nifB, encoding a radical SAM protein necessary for synthesizing a FeMo-co precursor and *nifX*, a carrier protein. For this, the evolution of Anf/Vnf/NifDKEN and NifB gene families was examined to infer the evolutionary origin of the Mo-nitrogenase and its evolutionary relationship to the alternative forms of the enzyme. These studies provide further evidence that Mo-nitrogenase emerged in the methanogenic archaea. In addition, the temporal relationships between the paralogous proteins necessary for nitrogen fixation and the synthesis of the active site cofactor and the biosynthesis of bacteriochlorophyll were examined. The light-independent (dark-operative) protochlorophyllide oxidoreductase (DPOR), encoded by *chlNB*/*bchNB* is necessary for bacteriochlorophyll synthesis in chlorophototrophs while anoxygenic phototrophs also encode the chlorophyllide reductase (*bchYZ*). BchN and BchZ are paralogs of Anf/Vnf/NifD while BchB and BchY are paralogs of Anf/Vnf/NifK. A series of gene duplications of an ancestral reductase is thought to have given rise to these paralogous proteins. A careful examination of these proteins has provided insight into the evolutionary linkages and temporal relationships between the emergence of photosynthesis relative to biological nitrogen fixation.

Materials and Methods

Phylogenetic Analysis

Representative homologs of Anf/Vnf/NifHDK and uncharacterized HDK (Table A1 of Appendix A) were compiled as previously described (*9*, *25*). Individual H, D, and K homologs were aligned using CLUSTALX (version 2.0.8) specifying the Gonnet 250 protein substitution matrix and default gap extension and opening penalties (*26*) as

previously described (*25*) with ChlLNB/BchLNB from *Anabaena variabilis* ATCC 29413 and *Chlorobium limicola* DSM 245 serving as outgroups. The individual alignment blocks were concatenated, subjected to evolutionary model prediction, and the phylogeny of each concatenated protein sequence evaluated using MrBayes (ver. 3.1.2) (*27*) and PhyML (ver. 3.0) (*28*) employing the WAG+I+G evolutionary model (See Appendix D for additional methods) as identified by ProtTest (version 2.0) (*29*). In phylogenetic reconstructions using MrBayes, tree topologies were sampled every 500 generations over 450,000 generations (after a burnin of 50,000) at likelihood stationarity and after convergence of two separate Markov chain Monte Carlo runs (average standard deviation of split frequencies <0.05). A consensus phylogenetic tree was projected from 1800 trees using FigTree (version 1.2.2) (http://tree.bio.ed.ac.uk/UH). One hundred bootstrap replicates were performed in phylogenetic reconstructions using PhyML. Matrices describing the Rao phylogenetic dissimilarity of concatenated HDK homologs, inferred using both MrBayes and PhyML, were generated using Phylocom (version 4.0.1) (*30*) (Fig. A1 in Appendix A).

Structural Analyses

The structures of the representative H, D, and K homologs selected for phylogenetic analysis were inferred using the CPH homology server for protein homology modeling (*31*) using the NifH (*32*), NifD (*33-35*), and NifK (*33-35*) crystal structures from *Azotobacter vinelandii* AvOP. The inferred structures for each H, D, and K homolog were imported into PyMol (ver. 1.4) (http://www.pymol.org/). The root-mean-square-deviations (RMSD) in the C^{α} positions were calculated for each individual

inferred H, D, and K homolog structure in relation to the other inferred H, D, or K homolog structures resulting in a pairwise matrix describing the structural RMSD (e.g., structural dissimilarity) for H, D, and K homologs. RMSDs generated for H, D, and K homologs, were normalized to compensate for differing HDK protein lengths, and the normalized H, D, and K matrices were then averaged to produce an HDK RMSD matrix for use in statistical analyses. PyMol was also used to generate images of sequence conservation in the active site cavity of Anf/Vnf/NifDK homologs.

Statistical Analyses

Mantel regressions of dissimilarity matrices were performed using XL Stat (ver. 2009.5.01). Ten thousand permutations employing two-tailed t-tests were used to determine the strength and significance of the relationships between dissimilarity matrices, respectively.

Vnf/Anf/NifDKEN and NifB Phylogenetic Analyses

BLASTp was used to compile all D, K, E, N, and B deduced amino acid sequences from genomic sequences using the DOE-IMG and the NCBI Genome Blast servers in July 2010 (Table A2 of Appendix A). Those which contained a full complement of DKENB were retained in the case of Vnf and Nif operons, or DKB in the case of Anf operons. Each individual locus was aligned using ClustalX (version 2.0.8) with the Gonnet 250 protein matrix and default gap extension and opening penalties (*26*). Each alignment was scrutinized and manually aligned using known catalytic residues (*36-38*). ProtTest (version 2.0) (*29*) was used to select WAG+I+G+F as the best-fit protein evolutionary model for individual Vnf/Anf/NifDKEN and NifB deduced amino

acid sequences and to determine the amino acid frequency (specified as 'F' in the model). The phylogeny of each deduced amino acid sequence was evaluated using MrBayes (version 3.1.2) (*39*, *40*) using the WAG evolutionary model with fixed (F) amino acid frequencies and gamma-distributed rate variation with a proportion of invariable sites (I+G). Tree topologies were sampled every 500 generations over 1.0×10^6 generations (after a burnin of 1.0×10^6) at likelihood stationarity and after convergence of two separate MCMC runs (average standard deviation of split frequencies less than 0.05). Taxa with D, K, E, and N deduced amino acid sequences that represented each of the major lineages were empirically selected (Table A2 in Appendix A) and these sequences were subjected to evolutionary analysis as described above (See Fig. A6 legend in Appendix A). A composite DKEN phylogram was constructed using the WAG evolutionary model with fixed amino acid frequencies (F) and gamma-distributed rate variation with a proportion of invariable sites (I+G). Tree topologies were sampled at likelihood stationarity as described above. The unrooted phylogram with collapsed clades (Fig. 2.1) was projected using TreeView (version 1.6.6) (*41*). The radial phylogram (Fig. 2.2) was projected with FigTree (version 1.2.2) (http://tree.bio.ed.ac.uk/UH). NifB deduced amino acid sequences from the taxa selected for the DKEN analysis were compiled and subjected to evolutionary analysis as described above, using the WAG+I+G+F substitution model. Tree topologies were sampled at likelihood stationarity as described above. The orientation of *nif* regulons from *Bradyrhizobium japonicum* and *Burkholderia phymatum* (Fig. 3.3B) have been reversed to conserve space. The genomic context of nitrogenase regulons was determined manually using the IMG Gene Neighborhood viewer.

Composite Vnf/Anf/NifDKEN and
BchNBYZ Phylogenetic Analyses

BLASTp was used to compile all BchNBYZ deduced amino acid sequences from genomic sequences using the DOE-IMG and the NCBI Genome Blast servers in November of 2009 (Table A3 of Appendix A). To simplify the analysis, bacteriochlorophyll biosynthesis sequences from eukarya were not considered in this analysis since previous studies indicate eukaryote bacteriochlorophyll biosynthesis sequences to be derived from bacterial sequences (*42*). Sequence alignment, evolutionary model selection, and phylogenetic analyses were performed as described above for each individual locus. Taxa representing the primary lineages for each individual BchNBYZ and Vnf/Anf/NifDKEN inferred protein sequence were used to construct a composite Vnf/Anf/NifDKEN/BchNBYZ phylogram. The radial phylogram (Table A3 of Appendix A) was projected with FigTree.

Evolutionary Rates Analysis

Estimated rates of substitution were converted to absolute ages for individual crown clades using three independent approaches including a data-driven penalized likelihood (PL) approach that allows for rate variation along branches of the phylogenetic tree (*43*), a molecular clock type approach that assumes a constant rate of evolution (*44*), and a nonparametric (NP) approach that seeks to minimize ancestor to descendent local rate changes (*45*), as implemented in version 1.71 of the program r8s (*43*). The 'VnfEN' lineage was used to root the tree. 'VnfEN' was chosen since it is present in both archaea and bacteria, it branches near the base of the tree, and is likely to have been present in the LUCA. While the limitations of rate smoothing and molecular clock approaches

including the non-clocklike behavior of proteins (*46*) and the bias towards overestimation of evolutionary time scales (*47*) are acknowledged and appreciated, we have used a range of time constraints to accommodate the various lines of evidence for the emergence of life (root age constraint) and for the emergence of oxygenic photosynthesis (oxygenic photosynthesis age constraint), and as previously mentioned, used three independent approaches to rate estimation. The various time constraints utilized include a fixed root age of either 3.4 Ga or 3.8 Ga and a fixed age for the emergence of oxygenic photosynthetic lineage of 2.5 Ga or 2.8 Ga. A root age of 3.8 Ga was chosen since isotopic evidence suggests that this may have been the earliest point in Earth's history that life emerged (*15, 16*). However, evidence in support of this date for life has been debated, and microfossil evidence substantiating this early calibration date have yet to be identified. Thus, a second calibration point of 3.4 Ga based on the earliest documented microfossil evidence for life on earth (*48*) was also included in the rate estimates. The age of the nodes demarcating the divergence of oxygenic and anoxygenic phototrophic lineages in the earliest branching of the two subclusters (C1) in each of the BchN and B lineages were calibrated to a conservative age of 2.5 Ga to reflect the timing of the Great Oxidation Event (GOE) (*5, 12, 13, 49-52*) as well as a calibration point of 2.8 Ga to accomodate additional evidence indicating that oxygenic phototrophs may have evolved by this time (*53-55*). Given that BchNB proteins from oxygenic phototrophs are nested among BchNB proteins derived from anoxygenic phototrophs in three of the four BchNB subclusters (Fig. A6 in Appendix 1), it is unlikely that mass extinction of oxygenic phototrophs has occurred. Thus, the crown clade ages in the earliest branching C1 subcluster of BchN and BchB reflect the minimum age estimates of the lineages and not

just the extant diversification. Importantly, each rate smoothing method (PL, molecular clock, NP) yielded similar estimated age mean and variances. Since derivations of age estimates from substitution rates centers on rate inconstancy, that the three independent methods produce similar results imply the sequences are evolving closer to rate constancy. The results of the PL method are presented (Figure 2.7), since this is a data-driven method of estimated substitution rates. The PL rate-smoothed chronogram (Table A3 of Appendix A) was projected with FigTree.

Results and Discussion

An Alternative Path for the Evolution of Biological Nitrogen Fixation

The limited availability of Mo under anoxic conditions led to the hypothesis that alternative forms of the nitrogenase enzyme lacking Mo emerged first. However, Bayesian and maximum-likelihood phylogenetic analyses of concatenated protein homologs of the required structural protein components (*nifHDK*, *vnfHDK*, and *anfHDK*) revealed that *nif*-encoded structural proteins H, D and K formed two distinct lineages, one of which branched at the base of the tree (Fig. 2.1). The basal NifHDK lineage was comprised of proteins derived solely from hydrogenotrophic methanogens. The second Nif lineage was comprised of more recently evolved Mo-dependent nitrogenase from both bacterial and methanogen genomes. Proteins of the Anf and Vnf nitrogenase are nested within the two Nif sublineages in a monophyletic lineage (Fig. 2.1). This indicates that Anf and Vnf were both derived from Nif and most likely resulted from gene duplication within the hydrogenotrophic methanogens and was not singly laterally

transferred.

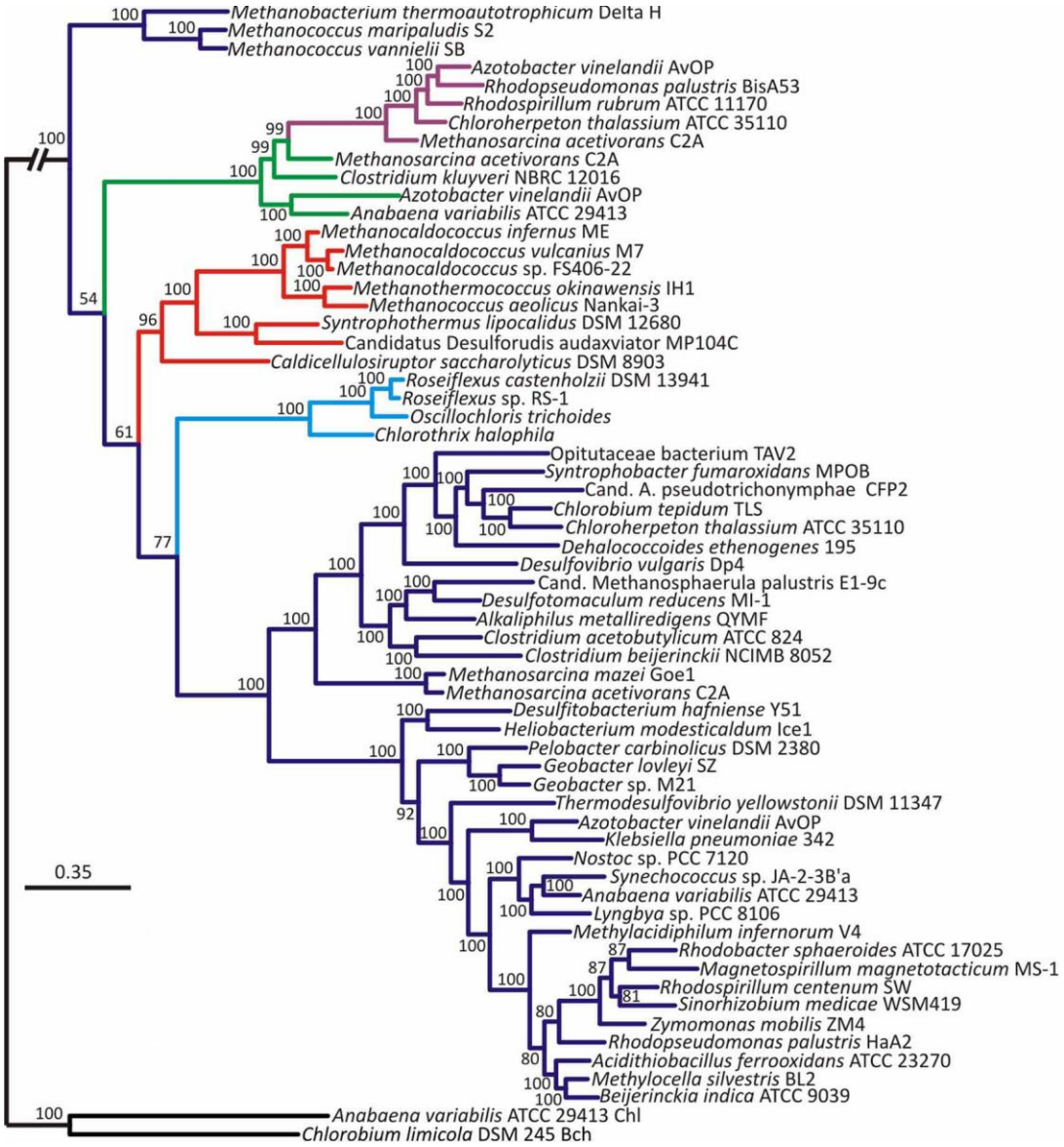


Figure 2.1. Bayesian inferred phylogenetic tree of concatenated HDK homologs and operon structure (See Fig. A1 for maximum-likelihood inferred tree in Appendix A). Posterior probabilities are indicated above or below nodes. Branches are colored dark blue (Mo-nitrogenase, Nif), green (V-nitrogenase, Vnf), purple (Fe-nitrogenase, Anf), red (uncharacterized nitrogenase), and light blue (uncharacterized homolog). The hash at the root was introduced to conserve space (See also Fig. A2 in Appendix A, maximum likelihood inferred phylogenetic tree).

This hypothesis is consistent with the observation that both of the alternative nitrogenases, Vnf and Anf, have yet to be identified in a genome that does not also encode Nif. In the monophyletic lineage containing the alternative nitrogenases, VnfHDK homologs nest the AnfHDK homologs, providing evidence that Anf derives from Vnf (Fig. 2.1). The observation that VnfHDK and AnfHDK from *Methanosarcinales* branch closely, coupled with the fact that these operons are located proximal in the genomes of these organisms, may suggest that *anf* is the result of a recent duplication of the *vnf* operon within the *Methanosarcinales* lineage. The acquisition of *vnf* within the *Methanosarcinales* lineage may have been the result of a lateral gene transfer (LGT) event with a firmicute, a finding that is consistent with the close spatial proximity noted between members of the *Methanosarcinales* and *Firmicutes* in a variety of anoxic environments (*56*) and with previous reports of LGT of individual genes and metabolic pathways between these two anaerobic lineages (*25, 57, 58*). Phylogenetic evidence indicates that *nif* may have been acquired in the *Methanosarcinales* via LGT (*8, 25*); however, it is unclear if that event predates the acquisition of *vnf* in this lineage and the subsequent duplication of *vnf* that resulted in *anf*. Biochemical evidence to date indicates that the biosynthesis of Anf and Vnf require *nif*-encoded gene products (*59-61*), suggesting the acquisition of *vnf* and the duplication of *vnf* that led to *anf* are most likely to postdate the acquisition of *nif* within this lineage.

Collectively, the evidence suggests that both Nif and Anf evolved in the methanogenic archaea, a guild of organisms which typically inhabit anoxic environments where Mo is in limited supply (*7*). Together with the fact that the expression of Anf and Vnf is tightly regulated by the availability of Mo and V (*61, 62*) and that the biosynthesis

of both Vnf and Anf examined to date require *nif*-encoded gene products for synthesis and acitivty, this set of observations suggests that the transient fluctuations in metal availability in anoxic environments may have been the impetus to incorporate new metals into the active site cluster of nitrogenase.

In addition to nitrogenase homologs in characterized diazotrophs, the phylogenetic history of HDK homologs encoded in genomes of organisms that have been shown to fix N_2 but where the metal content of the active site cofactors have not been determined was also examined. These 'uncharacterized nitrogenases' formed a monophyletic lineage in the analysis that branched after the Nif that were derived from hydrogenotrophic methanogens and the Anf/Vnf lineages (Fig. 2.1). This indicates that the uncharacterized nitrogenase emerged after both Nif and Vnf and possibly also predate Anf. The uncharacterized nitrogenase lineage is comprised of HDK proteins from methanogenic and methanotrophic archaea and the *Firmicutes*, all of which are strictly anaearobic taxa. In addition, a separate lineage comprised of uncharacterized HDK homologs found in the genomes of filamentous anoxygenic phototrophic bacteria branched after the Nif and Vnf lineages as well as the uncharacterized nitrogeases (Fig. 2.1). This indicates that these putative nitrogenases are the most recently evolved form of the enzyme although a physiological or biochemical role for these proteins has yet to be demonstrated.

Using homology modeling based on the well-characterized structures of *A. vinelandii* NifH and NifDK, the structures of the uncharacterized homologs were inferred. From the pairwise comparisons of the inferred protein structures, a matrix was generated from their structural dissimilarity. A significant and positive relationship

(Mantel $R^2 = 0.23$, $p < 0.01$) was noted between the regression of the pairwise structural dissimilarity matrix and a matrix describing the phylogenetic dissimilarity (Figure A3 of Appendix A). This data indicates that the structure of nitrogenase has changed significantly through time; however, residues that line the active site pocket where the cofactor binds are well-conserved, according to primary sequence alignments and the structures inferred from homology modeling (Figures A4 & A5 of Appendix A). This suggests that once the active site cavity evolved to house the metal cofactor necessary for dinitrogen reduction, the residues in the cluster binding cavity and the cluster-coordinating ligands were maintained, regardless of the metal composition of the cofactor.

We examined phylogenetic and structural relationships among proteins that are evolutionarily related to nitrogenase, including those required to biosynthesize bacteriochlorophyll (BchN) (*63, 64*) and those that have been proposed to catalyze an analogous reaction in Ni porphyrin F_{430} biosynthesis (NflD) (*65*). Phylogenetic reconstruction of these proteins, NflD, BchN and Anf/Vnf/NifD revealed three distinct lineages for each protein homolog (Fig. 2.2). NflD proteins formed a lineage that bisects the lineage of Anf/Vnf/NifD and the lineage of BchN. This analysis indicates NflD is ancestral to Anf/Vnf/NifD and BchN, a finding that is consistent with a previous phylogenetic analysis of concatenated NflHD, Anf/Vnf/NifHD and BchLN that indicated the Nfl proteins are ancestral (*11, 19*). Intriguingly, NflD proteins share little sequence conservation with the active site cavity of Anf/Vnf/NifD and BchN. Likewise, the cofactor coordinating ligands of the nitrogenase homolog proteins Anf/Vnf/NifD are not

conserved in BchN and the recently solved crystal structure of BchNB (*66, 67*) indicates an open cavity in the protein necessary for bacterochlorophyll biosynthesis where the complex metal cofactor binds in nitrogenase homologs. The derived states, Anf/Vnf/NifDK and BchNB, have maintained similar structural architecture to the inferred ancestral state (NflD); however, the cavity residues necessary for cofactor (metal cofactor for nitrogenase) or substrate binding (protochlorophyllide binding to BchNB) have been fine-tuned to bind the cofactor or substrate as the paralogs have diversified.

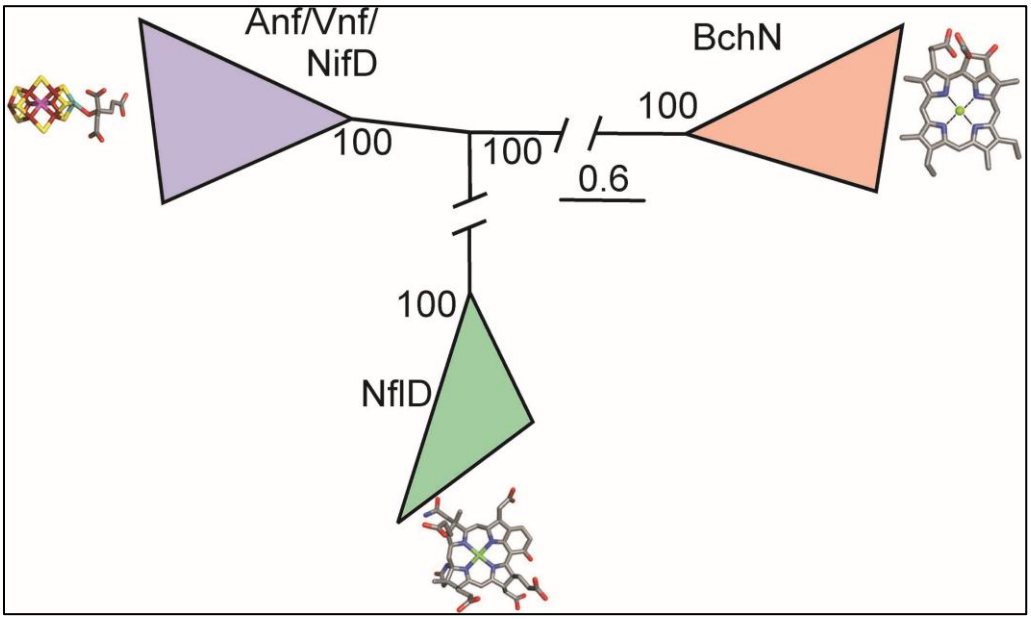


Figure 2.2. Bayesian inferred phylogenetic reconstruction of Anf/Vnf/NifD, BchN, and NflD proteins. The putative substrates and cofactors for each protein lineage are indicated below each respective clade. Posterior probabilities for each collapsed node are indicated. Nodes have been collapsed and hashes introduced to conserve space.

These studies have led to a hypothesis for the emergence of nitrogenase in which a gene that encoded for an ancestral protein complex with a cavity similar to that observed in the inferred structure of NflD duplicated, leading the evolutionary precursors of Anf/Vnf/NifD and BchN (Fig. 2.3).

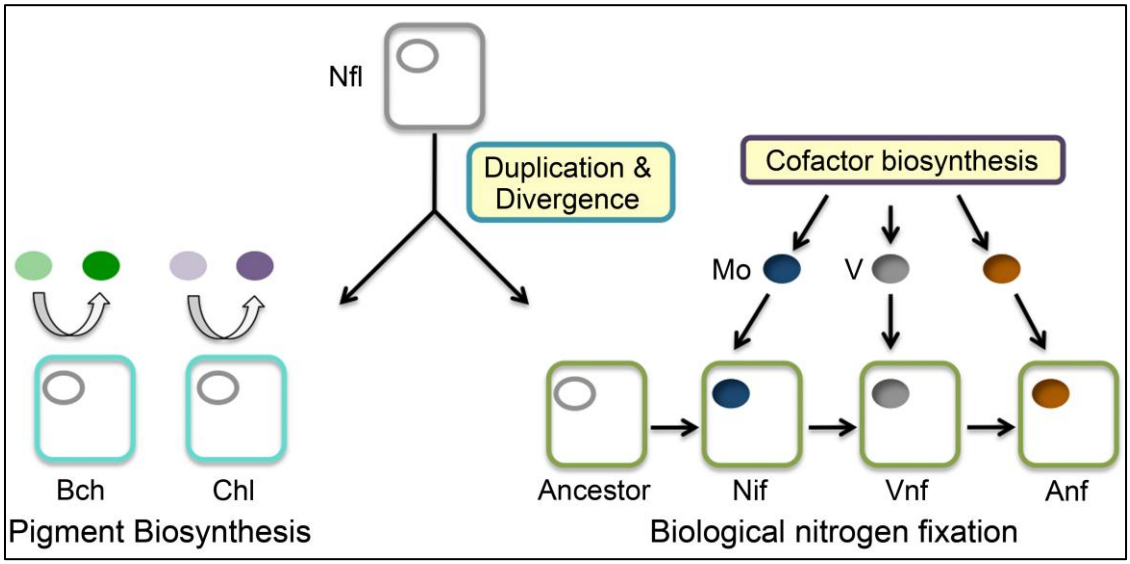


Figure 2.3. Model depicting the divergence of nitrogenase (NifD) and protochlorophyllide reductase (ChlN/BchN) from a NflD ancestor. The stepwise evolution of cofactor biosynthesis leading to the acquisition of metal specificity in the covalently bound active site metallocluster, where Mo acquisition and Mo-nitrogenase predates V acquisition and V-nitrogenase, and V acquisition predates Fe-only nitrogenase. ChlN/BchN bind substrates in their active site cavities non-covalently and release these substrates following reduction (*68*). Abbreviations: Mo, molybdenum; V, vanadium; Nif, Mo-dependent nitrogenase; Vnf, V-dependent nitrogenase; Anf, Fe-only nitrogenase; Bch, BchN protein involved in bacteriochlorophyll biosynthesis; Chl, ChlN protein involved in chlorophyll biosynthesis.

Metals or metal clusters were bound serendipitously in the cavity of the ancestral precursor in a nonspecific manner, resulting in an enzyme complex with altered and perhaps preferential reactivity such as N_2 reduction. Limited fixed nitrogen on early Earth would have provided the selective pressure to maintain the use of the metal cofactor with preferential activity in addition to the recruitment of genes and associated gene products to improve the catalytic ability of the enzyme through the modification of the metal cofactor. In addition, the binding site cavity also evolved to bind the active site metal cofactor, FeMo-co in the Mo-nitrogenase, to fine-tune the structural determinants for nitrogenase catalysis. This hypothesis is consistent with the end-to-end dimensions of

FeMo-co which is similar to both bacteriochlorophyll, the substrate for BchNB and F_{430}, the presumed substrate for NflD, suggesting the size and dimensions of the cofactor were constrained by the structure and active site cavity of the ancestor. Importantly, as-isolated FeMo-co is not reactive toward N_2 reduction and is only active upon incorporation into NifDK, implying that the hypothesized stepwise evolution of nitrogenase may be the only mechanism by which the biochemical pathway for cofactor biosynthesis could have evolved as a result of selective pressure due to fixed nitrogen limitation.

A Late Methanogenic Origin for Molybdenum-dependent Nitrogenase

Origin of Nif

Several of the genes involved in the biosynthesis of FeMo-co have arisen from gene duplication and gene fusion events, based on evidence to date. NifEN, the molecular scaffold where the FeMo-co is assembled encoded by *nifE* and *nifN*, was derived from a paralogous gene duplication event of *nifD* and *nifK* (Fig. 2.4A). As reported previously, nitrogenase regulons containing D, K, E and N were identified in the genomes of some of the methanogenic archaea as well as a variety of phylogenetically distinct bacteria (Fig. 2.4B). Also, homologs of these loci were absent from all other lineages of the archaea (*Crenarchaea, Korarchaea,* and *Nanoarchaea*) and early branching members of the bacteria (*Thermotogae*, and *Deinococcus*/*Thermus*). Interestingly, nitrogenase homologs were identified in the genome of one *Aquificae, Thermocrinis albus* DSM, one of the deepest branching bacterial lineages. However, these regulons were incomplete, containing only homologs of *nifD, nifK* and *nifN*. In addition, the branching order of

these individual proteins in *T. albus* suggests they were acquired by LGT with either a firmicute or a proteobacterium. As previously observed, the distribution of *anf* and *vnf* operons in genomes was far more restricted than the distribution of *nif*. Anf operons were only identified in genomes that also encoded for Nif, a finding that is consistent with the alternative nitrogenases requiring *nif*-encoded E and N homologs. Similarly, Vnf operons were also only identified in the genomes of organisms that also encoded Nif.

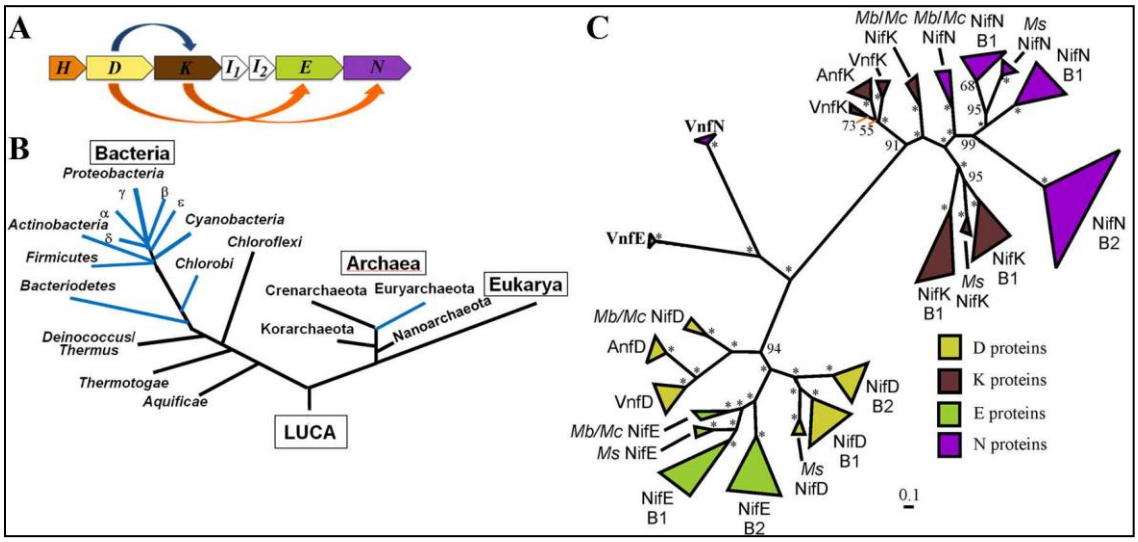


Figure 2.4. (A) Nitrogenase regulon from *Klebsiella pneumoniae* with the blue arrow indicating the duplication event of *nifD* yielding *nifK* and the orange arrows indicating the in-tandem duplication event of *nifDK* yielding *nifEN*. (B) Universal tree indicating lineages where full complements of nitrogenase proteins (HDKENB) were found (blue lineages). (C) Bayesian consensus phylogenetic tree of Nif/Vnf/AnfD, K, E, and N proteins from 40 taxa that represent the primary lineages for each of the 4 loci (C). The width and depth of collapsed clades proportionally reflect the number and the diversity of sequences in that clade, respectively. *Methanobacteriales*/*Methanococcales* proteins are abbreviated "*Mb*/*Mc"* and *Methanosarcinales* proteins are abbreviated by "*Ms"*. All lineages labeled as Nif without a *Mb*/*Mc* or *Ms* designation contain only bacterial sequences. Asterisks indicate 100% posterior probability support; bar equals 1 substitution per 10 sequence positions.

Deduced amino acid sequences of D, K, E, and N provide a root for each other in evolutionary reconstructions since the genes encoding these proteins arose from a gene

duplication event. This observation was used to examine the evolutionary history of the Mo-nitrogenase and to provide further evidence that Mo-nitrogenase was not associated with the LUCA and most likely arose after the split of bacteria and archaea. For this analysis, the evolutionary history of the deduced amino acid sequences of DKEN was reconstructed and the branching order of each homolog was examined for each individual sequence. The resulting phylogram revealed well-supported lineages that corresponded to metal content (Fig. 2.4C). This analysis revealed that that NifD deduced amino acid sequences were paraphyletic with respect to NifE (Fig. 2.4C). The two lineages of NifD sequences contained only archaeal sequences while the lineage was comprised of both archaeal and bacterial sequences.

The deduced amino acid sequences of NifK were also paraphyletic with respect to NifN with one comprised of archaeal NifK and the other both bacterial and archaeal NifK (Fig. 2.4C). Importantly, the NifE and NifN lineages were both monophyletic and nested within NifD and NifK, respectively. The fact the NifE nests within NifD and NifN within NifK provides strong support for an evolutionary scenario where duplication of an ancestral *nifDK* resulted in *nifEN*. This duplication event to form the molecular scaffold protein NifEN presumably enabled further evolution of the FeMo-co precursor, resulting in a reductase cofactor with superior reactivity than that synthesized in the absence of NifEN. In this analysis, the Mo-nitrogenase is defined as having *nifEN*, indicating that this form of the nitrogenase enzyme was not present in the LUCA.

Methanogens of the order *Methanococcales* (*Methanococcus maripaludis* and *Methanococcus vannielli*; denoted by "*Mc*") and *Methanobacteriales* (*Methanothermobacter thermoautotrophicus*; denoted by "*Mb*") form distinct groups in

all individual Nif proteins D, K, E and N, that diverged before all other Nif proteins, including those from the *Methanosarcinales* (*Methanosarcina barkeri* and *Methanosarcina acetivorans*; denoted by "*Ms*") (Fig. 2.4C and Fig. 2.5). This finding is consistent with phylogenetic analysis of both nifH and 16S rRNA genes. In these analyses the more metabolically versatile *Methanosarcinales* branch later than genes from members of the *Methancoccales* and *Methanobacteriales*, which utilize H_2 to reduce CO_2 to CH_4. The topology of NifD and NifK lineages, where *Methanococcales*/*Methanobacteriales* and *Methanosarcinales*/bacterial sequences are paraphyletic with respect to NifE and NifN (Fig. 2.4C, Fig. 2.5), suggests that the duplication of *nifDK* resulting in *nifEN* likely occurred in an ancestor of the hydrogenotrophic methanogens. Among the bacterial taxa, two primary lineages of NifDKEN are apparent in the phylogenetic reconstruction. One bacterial lineage is comprised of sequences from strictly anaerobic taxa (*Firmicutes, Proteobacteria,* Green Non-Sulfur bacteria, and Green Sulfur bacteria) and these form a sister clade to proteins from the *Methanosarcinales* in each NifDE lineage (Fig. 2.4C, Fig. 2.5), but interestingly are paraphyletic with respect to *Methanosarcinales* NifEN sequences. The other lineage contains sequences from both aerobic and anaerobic taxa. In the second cluster, the strict anaerobes *Heliobacterium modesticaldum* and *Desulfitobacterium hafniense* branch as the base.

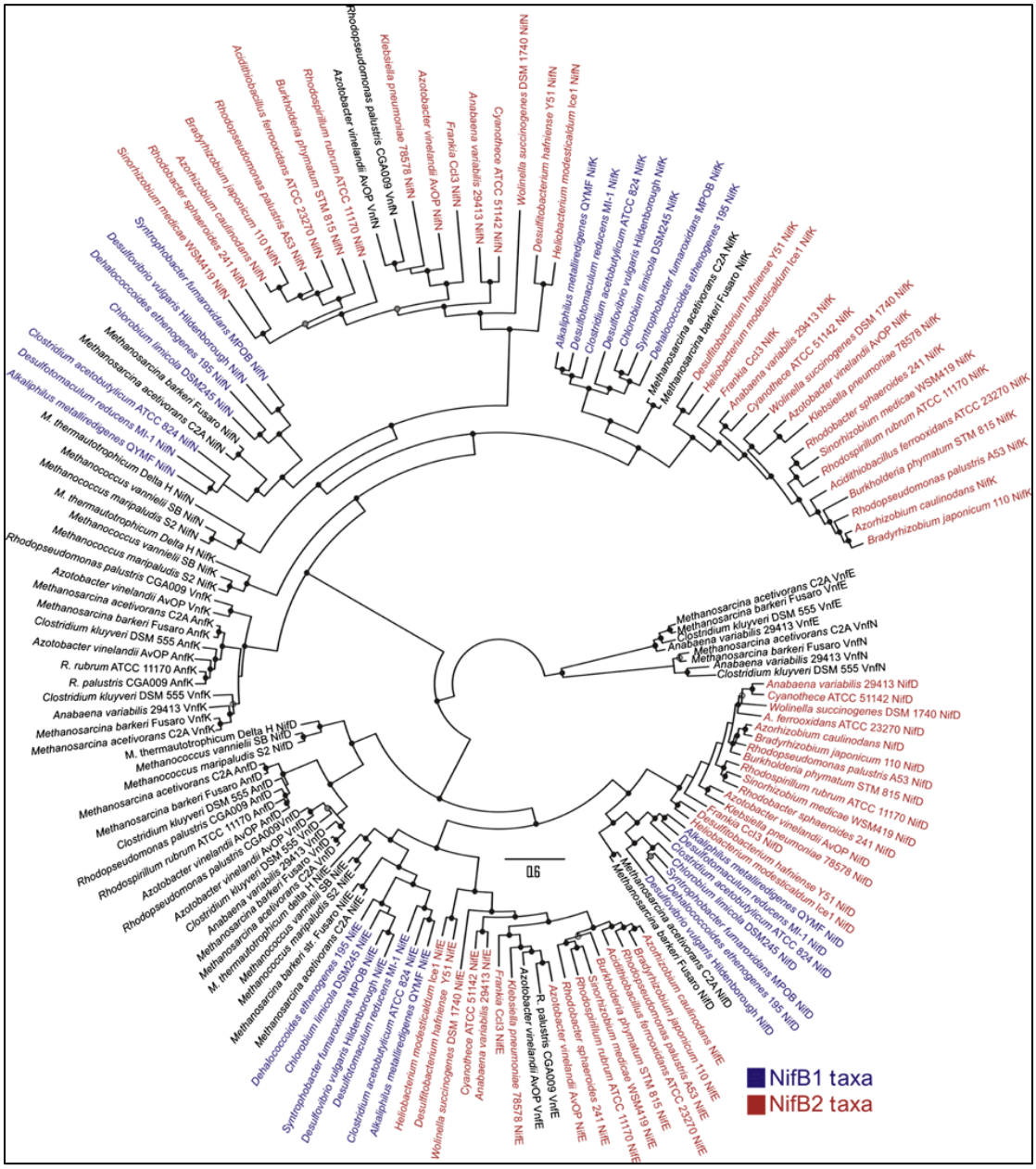


Figure 2.5. Bayesian consensus phylogenetic tree of Nif/Vnf/AnfD, K, E, and N proteins from 40 taxa that represent the primary lineages for each of the 4 loci. Black circles at nodes denote >90% posterior probability (PP), gray circles at nodes denote >80% PP, open circles denote >70% PP, and no symbol denotes 50 to 70% PP. Nodes with <50% PP were collapsed; bar equals 6 substitutions per 10 sequence positions (See also, Figure A6 in Appendix A).

The branching of these lineages with respect to the lineages containing sequences from the methanogenic archaea suggest a role for a LGT or multiple LGT events in the

acquisition of Nif in bacteria, with transfer from a methanogen to an ancestral member of the *Firmicutes*. The likelihood of this hypothesis is bolstered by the close spatial proximity observed for members of the *Methanosarcinales* and *Firmicutes* in a variety of extant anoxic environments as well as the number of reports of LGT events between two anaerobic lineages. The origin and early proliferation of Mo-nitrogenase in strict anaerobic lineages of both bacteria and archaea is also consistent with the extreme oxygen sensitivity of the nitrogenase complex. In addition to the structural proteins NifDK and the molecular scaffold NifEN, the phylogeny and genetic structure of NifB, a radical SAM protein necessary for biosynthesis of FeMo-co, was also examined from all taxa that contain a full complement of NifDKEN. These analyses were performed to further investigate that likelihood of an archaeal origin of Mo-nitrogenase. Alignment of NifB sequences from taxa containing NifDKEN revealed that the methanogenic archaea and a few anaerobic bacterial diazotrophs contained only the SAM domain and lacked the C-terminal 'NifX' domain (Fig. 2.6). Phylogenetic reconstruction of the SAM domain of each of the NifB sequences indicated that all NifB sequences lacking the NifX domain diverged early with respect the NifB sequences that contain the fused NifX domain (Fig. 2.6).

Of these early branching sequences, NifB from both *Methanococcales* and *Methanobacteriales* formed well-supported and early branching lineages, indicating that these NifB sequences evolved early with respect to NifB from the *Methanosarcinales* and *Bacteria*. A sister group of NifB from members of the *Firmicutes* with members of the *Methanosarcinales* further supports the hypothesis that a LGT event or events resulted in

acquisition of Mo-nitrogenase in the bacteria via an interaction between an ancestral member of the *Firmicutes* and an ancestral member of the *Methansarcinales.*

The genetic structure of *nif* regulons containing the full complement of *nifDKEN* was examined to compare regulons that contained *nifB* genes encoding for fused 'SAM-NifX' domains with those that contained *nifB* genes encoding for only the SAM domain.

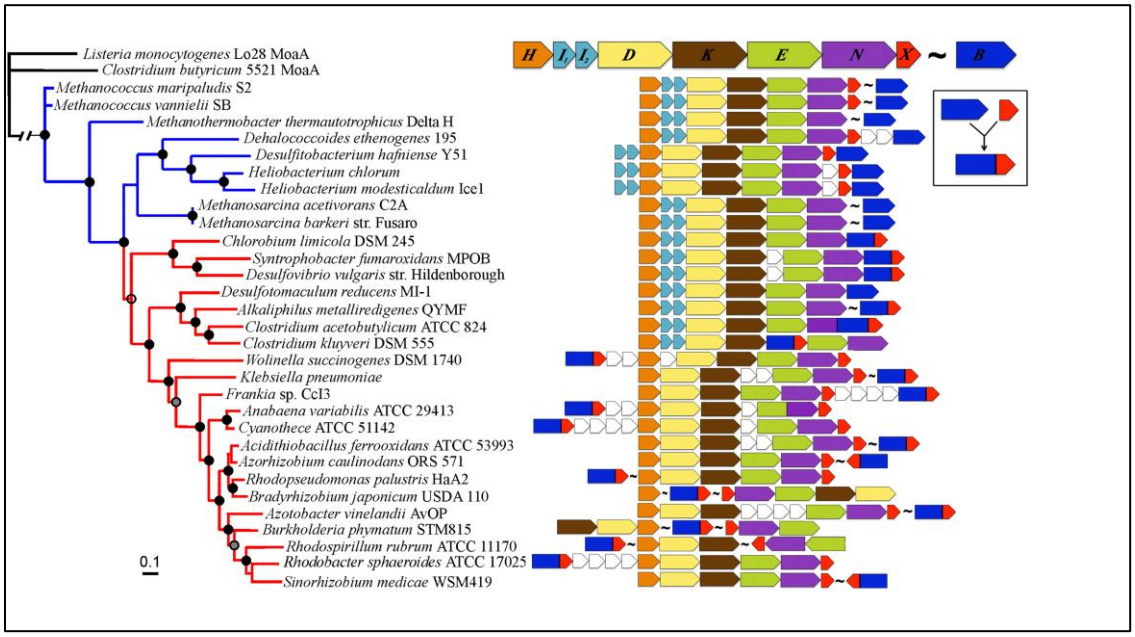


Figure 2.6. Phylogenetic tree of the radical SAM domain of NifB from 30 taxa that contain NifDKEN rooted with the molybdenum biosynthesis protein (MoaA) is shown in the left panel. Taxa were selected to represent the major lineages of NifB. Lineages highlighted in blue denote taxa with NifB proteins which contain only the SAM domain and lineages highlighted in red denote fused 'SAM-NifX' proteins. Asterisks denote 100% posterior probability support; bar, 1 substitution per 10 sequence positions. Genetic structure of the *nif* regulons for taxa presented in phylogenic tree are shown in the right panel. A gene locus color key is indicated at the top of the panel and an illustration of the NifB-NifX fusion event is inset. The orientation of *nif* regulons from *Bradyrhizobium japonicum* and *Burkholderia phymatum* have been reversed to conserve space.

In contrast to *nif* regulons from genomes of organisms containing genes encoding for fused NifB-NifX proteins, the gene order is generally more conserved in organisms containing the early branching NifB sequences lacking the 'NifX' fusion domain (Fig.

2.6, right panel). The synteny of *nif* regulons from a diversity of methanogenic genomes is consistent with the hypothesis that Mo-nitrogenase emerged in the hydrogenotrophic methanogen lineage. Moreover, the lack of the fused NifB-NifX in anaerobic members of the *Firmicutes*, as well as similar operon structures in these organisms supports the hypothesis that *nif* was acquired in the bacterial domain via LGT from a methanogen to a firmicute.

Temporal Relationship Between the Emergence of Nif and Oxygenic Photosynthesis

Phylogenetic reconstruction of the increasing number of available genomes has provided significant support for the hypothesis that Mo-nitrogenase emerged in a methanogenic archaea. It has been suggested that the proliferation of Mo-nitrogenase was precipitated by the rise in marine Mo concentration that accompanied the Great Oxidation Event (GOE), the point at O_2 appeared in the atmosphere approximately 2.4 Ga (*69*). However, it remains unclear whether Mo-nitrogenase emerged prior to the GOE when bioavailable levels of Mo were low or if it emerged after the GOE when Mo became much more available. In order to examine the timing of the emergence of Nif in the genomic record relative to the geologic record, the evolutionary relationships of proteins involved in nitrogen fixation (Nif) and bacteriochlorophyll biosynthesis (Bch) were examined. This analysis elucidated whether or not Mo-nitrogenase, specifically the duplication of *nifDK* to *nifEN*, occurred prior to the divergence of oxygenic and anoxygenic phototrophs and when Mo was available in very limited supply or after when environmental oxygenation was widespread. To determine the relative plausibility of

each of these scenarios, the evolutionary history of representative Vnf/Anf/NifDKEN and BchNBYZ deduced amino acid sequences was reconstructed.

The Vnf/Anf/NifDKEN and BchNBYZ phylogram reveals well-supported and distinct sublineages that cluster within either the nitrogenase or Bch stem lineages of the tree (Fig. A7 & A8 in Appendix A) Within the BchNB subclusters, anoxygenic phototrophs branch early with respect to oxygenic phototrophs, supporting other analyses of proteins involved in photosynthesis that indicate anoxygenic photosynthesis emerged prior to oxygenic photosynthesis (Figure A7 in Appendix A). While these analyses indicate an evolutionary trajectory of chlorophotrophs, age estimates combined with geologic record may provide insight into which enzymatic function first emerged.

Three independent rate smoothing approaches (PL, rate constancy, and NP) were employed to compare the ages of the four crown clades corresponding to BchN, BchB, NifDE and NifKN. All three approaches yielded similar age estimates. Using the PL algorithm with the root age set to 3.5 Ga and the age of the emergence of oxygenic photosynthesis set to 2.5 Ga, the estimated age of the Anf/Vnf/NifDE crown clade was 2.07 ± 0.10 Ga and the estimated age of the Vnf/Anf/NifKN crown clade was 1.84 ± 0.20 Ga (Fig. 2.7. The estimated age of the NifDE crown clade (2.07 ± 0.10 Ga) and the NifKN crown clade (1.84 ± 0.20 Ga) (Fig. 2.7, Fig. A9 in Appendix A) were similar, providing further support for an in tandem duplication of *nifDK* to *nifEN.* After incorporating the age estimates from all substitution rate estimate iterations using the various time constraints, the boundaries on the timing of the duplication of *nifDK* to *nifEN* can be estimated to be 1.5 to 2.2 Ga.

The ages of NifDE and NifKN clades are each estimated at about 1.5 to 2.2 Ga, suggesting that the timing of the duplication of *nifDK* to *nifEN* and the ability to incorporate Mo into the nitrogenase co-factor occurred prior the emergence of oxygenic photosynthesis at ~ 2.5 to 2.8 Ga (*5, 12, 13*).

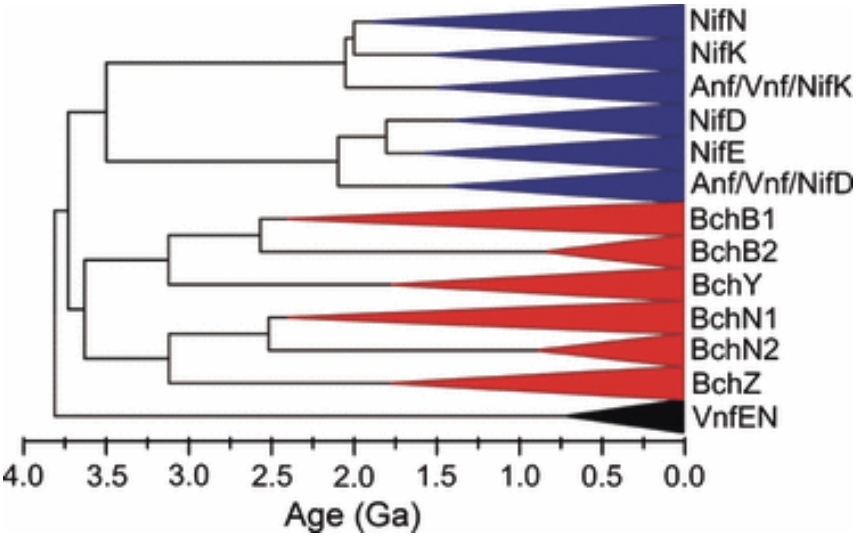


Figure 2.7. Penalized-likelihood rate-smoothed chronogram of representative sequences for Vnf/Anf/NifDKEN and BchNBYZ rooted with representatives from the VnfEN lineage. Nitrogenase and Bch protein lineages are shaded in gray and black, respectively. An uncollapsed version of the rate-smoothed chronogram is presented in Figure A3 of Appendix A.

An origin for Mo-nitrogenase that follows the emergence of oxygenic photosynthesis is consistent with the distribution of *nifDKEN* in extant genomes and with phylogenetic reconstructions of NifDKEN, which collectively suggest that Nif evolved relatively recently and after the split between bacteria and archaea. The emergence for Mo-nitrogenase at ~1.5 to 2.2 Ga would be consistent with the hypothesized nitrogen crisis estimated at 2.2 Ga (*70*). Furthermore, geochemical measurements indicate a rise in Mo concentrations in the oceans preceding this time at ~2.5 Ga (*6*), due to increased

weathering of continental sulfides in the presence of an increasingly oxidizing environment. (*71*) The emergence of Mo-nitrogenase prior to the GOE could have been due to an increase in Mo availability in response to the gradual oxygenation of the oceans that would have occurred prior to the GOE. The increase in Mo during this time could have provided the selective pressure necessary to evolve the catalytically superior Mo-containing form of the active site metal cluster and thus the biosynthetic machinery necessary to synthesize this cofactor. The phylogenetic data however, suggest the origin of Mo-nitrogenase in an anoxic environment after the rise of O_2. The observation that the earliest branching extant Mo-nitrogenases are from hydrogenotrophic methanogens, which are strict anaerobes that also persist in environments that are normally under reducing conditions and thus restrict the bioavailability of Mo lead to a hypothesis that Nif emerged during the earliest stages of the GOE. Such conditions would have been common before the oceans became fully oxygenated with episodic availability of Mo. During the early Proterozoic (~1.5 to 2.2 Ga) when Mo nitrogenase is thought to have originated, ocean basins were commonly poor in SO_4^{2-} (*13*), especially in waters extending greater than 100 km beyond the continental shelf (*72*). In the absence of significant concentrations of SO_4^{2-} (<60 µM), hydrogenotrophic methanogens can outcompete sulfate-reducing bacteria (SRB) for the growth substrate H_2 (*73, 74*). Therefore, the effects of low concentrations of sulfate in a redox stratified ocean basin may have promoted the emergence of Mo-nitrogenase within the hydrogenotrophic methanogens. Decreased SRB activity would preclude significant contributions of biologically-sourced H_2S capable of precipitating episodic flux of Mo from overlaying

waters (*7*), thereby creating a niche that would presumably have been prime for hydrogenotrophic methanogens to persist and evolve Mo-nitrogenase.

Conclusions

In summary, the Mo-nitrogenase in extant biology is not likely to be the first nitrogenase associated with early life on Earth, a finding that is in line with the evidence supported by geochemistry (*75, 76*). However, in contrast with what has been proposed based on geochemical evidence (*8, 75, 76*), these results indicate that alternative nitrogenases (V- and Fe-only forms) are not ancestors of the Mo-nitrogenase but rather are derived from Mo-nitrogenase. The common ancestor of Nif/Vnf/Anf, Bch/Chl, and Nfl presumably had a cavity capable of binding certain porphyrins and/or metal cluster fragments that has evoloved to bind the specific substrates or active site metal clusters observed in the extant enzymes. The nature of the ancestral nitrogenase enzyme and its associated bound metal cluster was likely controlled by the selective pressure imposed by fixed nitrogen limitation in combination with local environmental metal availability until the GOE when Mo became sufficiently bioavailable (*75, 76*) and the most biochemically favorable form of nitrogenase (Nif) emerged that is reflective modern forms of the enzyme. These results reveal a new paradigm for the evolution of biological nitrogen fixation in which the Mo-nitrogenase emerged prior to the alternative forms and provide key insights into the manner in which early life forms might have exploited the reactivity of their mineral environment prior to evolving the refined complex metalloenzymes observed today.

Key events (e.g., gene duplications, gene fusions) in the evolution of enzymes that are required to biosynthesize the FeMo active site co-factor of Mo-nitrogenase generate new insight into the evolution of biological nitrogen fixation. The temporal relationship of Mo-nitrogenase evolution with the evolution of oxygenic photosynthesis was examined to investigate the hypothesis that Nif was not associated with the LUCA and in fact emerged in the methanogenic archaea. Phylogenetic reconstruction of Anf/Vnf/NifDKEN sequences indicate that NifEN are nested among NifDK sequences, suggesting that *nifDK* duplicated and gave rise to *nifEN*. The phylogenetic branching order of NifDKEN deduced amino acid sequences suggests that the duplication of *nifDK* to *nifEN* most likely occurred in an ancestor of the hydrogenotrophic methanogens, with acquisition in the bacterial domain via LGT between an ancestor of the methanogenic archaea and an ancestor of the *Firmicutes*. A comparison of substitution rates in Anf/Vnf/NifDKEN deduced amino acid sequences and substitution rates of BchNBYZ deduced amino acid sequences suggests that Nif emerged ~ 1.5 to 2.2 Ga, after the origin of oxygenic photosynthesis and the widespread oxygenation of the biosphere. Given the evidence for the emergence of Nif in an ancestor of a hydogenotrophic methanogen with subsequent acquisition in bacteria via LGT to a member of the *Firmicutes*; it is likely that the origin and early proliferation of Nif occurred in an anoxic environment where Mo is at least episodically available. Collectively, these results provide evidence that Mo-nitrogenase emerged in an ancestor of hydrogenotrophic methanogens, prior to either of the alternative forms of the enzyme. Phylogenetic reconstruction and substitution rate estimates provide a hypothesis for the origin of Mo-nitrogenase in a redox-stratified ocean where mildly-oxygenated Mo-containing surface waters overlay deeper waters that

remain depleted in O_2 and H_2S. Such an environment would create a niche ideal for hydrogenotrophic methanogens and the selective pressure to evolve the Mo-nitrogenase.

Acknowledgments

We are indebted to Robert Blankenship, Donald Bryant, and Luis Rubio for stimulating discussions and/or critical evaluations of the manuscript. We thank Aubrey Zerkle and two anonymous reviewers whose comments and suggestions greatly improved this manuscript. This work was supported by the NASA Astrobiology Institute (J.W.P and A.D.A.) and the National Science Foundation (S.R.M.). E.S.B. was supported by a NASA Astrobiology Institute postdoctoral fellowship and T.L.H. is supported by a National Science Foundation – Integrative Graduate Education and Research Training fellowship.

Supporting Information Available: Additional figures and tables for this study are available free of charge via the Internet at http://onlinelibrary.wiley.com/doi/10.1111/j.1472-4669.2011.00278.x/suppinfo and http://www.frontiersin.org/microbiological_chemistry/10.3389/fmicb.2011.00205/full#. The relevant data is located in Appendix A.

Literature Cited

1. Yung, Y. L., and McElroy, M. B. (1979) Fixation of nitrogen in the prebiotic atmosphere, *Science 203*, 1002-1004.

2. Kasting, J. F. (1981) Limits on Oxygen Concentration in the Prebiological Atmosphere and the Rate of Abiotic Fixation of Nitrogen, *Journal of Geophysical Research 86*, 1147 - 1158.

3. Rubio, L. M., and Ludden, P. W. (2008) Biosynthesis of the iron-molybdenum cofactor of nitrogenase, *Annu. Rev. Microbiol. 62*, 93-111.

4. Zehr, J. P., Jenkins, B. D., Short, S. M., and Steward, G. F. (2003) Nitrogenase gene diversity and microbial community structure: a cross-system comparison, *Environ Microbiol 5*, 539-554.

5. Anbar, A. D., Duan, Y., Lyons, T. W., Arnold, G. L., Kendall, B., Creaser, R. A., Kaufman, A. J., Gordon, G. W., Scott, C., Garvin, J., and Buick, R. (2007) A whiff of oxygen before the great oxidation event?, *Science 317*, 1903-1906.

6. Scott, C., Lyons, T. W., Bekker, A., Shen, Y., Poulton, S. W., Chu, X., and Anbar, A. D. (2008) Tracing stepwise oxygenation of the Proterozoic ocean, *Nature 452*, 456-459.

7. Helz, G. R., Miller, C. V., Charnock, J. M., Mosselmans, J. F. W., Pattrick, R. A. D., Garner, C. D., and Vaughan, D. J. (1996) Mechanism of molybdenum removal from the sea and its concentration in black shales: EXAFS evidence, *Geochim. Cosmochim. Acta 60*, 3631-3642.

8. Raymond, J., Siefert, J. L., Staples, C. R., and Blankenship, R. E. (2004) The natural history of nitrogen fixation, *Mol. Biol. Evol 21*, 541-554.

9. Soboh, B., Boyd, E. S., Zhao, D., Peters, J. W., and Rubio, L. M. (2010) Substrate specificity and evolutionary implications of a NifDK enzyme carrying NifB-co at its active site, *FEBS Letters 584*, 1487-1492.

10. Leigh, J. A. (2000) Nitrogen fixation in methanogens: The archaeal perspective, *Curr. Issues Mol. Biol. 2*, 125-131.

11. Raymond, J., Siefert, J. L., Staples, C. R., and Blankenship, R. E. (2004) The natural history of nitrogen fixation, *Mol Biol Evol 21*, 541-554.

12. Holland, H. D. (2002) Volcanic gases, black smokers, and the Great Oxidation Event. , *Geochimica et cosmochimica acta 66*, 3811–3826.

13. Canfield, D. E., Habicht, K. S., and Thamdrup, B. (2000) The Archean Sulfur Cycle and the Early History of Atmospheric Oxygen, *Science 288*, 658-661.

14. Penny, D., and Poole, A. (1999) The nature of the last universal common ancestor, *Curr. Opin. Gen. Dev. 9*, 672–677.

15. Mojzsis, S. J., Arrhenius, G., McKeegan, K. D., Harrison, T. M., Nutman, A. P., and Friend, C. R. L. (1996) Evidence for life on Earth before 3,800 million years ago, *Nature 384*, 55-59.

16. Parsons, I., Lee, M. R., and Smith, J. V. (1998) Biochemical evolution II: Origin of life in tubular microstructures on weathered feldspar surfaces, *Proc. Natl. Acad. Sci. USA 95*, 15173–15176.

17. Boone, D. R., Whitman, W. B., and Rouvière, P., (Eds.) (1993) *Methanogenesis*, Chapman & Hall, New York & London.

18. Thauer, R. K., Kaster, A.-K., Seedorf, H., Buckel, W., and Hedderich, R. (2008) Methanogenic archaea: ecologically relevant differences in energy conservation, *Nat. Rev. Microbiol. 6*, 579-591.

19. Staples, C. R., Lahiri, S., Raymond, J., Von Herbulis, L., Mukhophadhyay, B., and Blankenship, R. E. (2007) Expression and association of group IV nitrogenase NifD and NifH homologs in the non-nitrogen-fixing archaeon Methanocaldococcus jannaschii, *J Bacteriol 189*, 7392-7398.

20. Rubio, L. M., and Ludden, P. W. (2008) Biosynthesis of the iron-molybdenum cofactor of nitrogenase, *Annu Rev Microbiol 62*, 93-111.

21. Joerger, R. D., and Bishop, P. E. (1988) Bacterial alternative nitrogen fixation systems, *Crit Rev Microbiol 16*, 1-14.

22. Soboh, B., Boyd, E. S., Zhao, D., Peters, J. W., and Rubio, L. M. (2011) Substrate specificity and evolutionary implications of a NifDK enzyme carrying NifB-co at its active site, *FEBS Lett 584*, 1487-1492.

23. Fani, R., Gallo, R., and Lio, P. (2000) Molecular evolution of nitrogen fixation: the evolutionary history of the nifD, nifK, nifE, and nifN genes, *J Mol Evol 51*, 1-11.

24. Curatti, L., Ludden, P. W., and Rubio, L. M. (2006) NifB-dependent in vitro synthesis of the iron-molybdenum cofactor of nitrogenase, *Proc Natl Acad Sci U S A 103*, 5297-5301.

25. Boyd, E. S., Anbar, A. D., Miller, S., Hamilton, T. L., Lavin, M., and Peters, J. W. (2011) A late methanogen origin for molybdenum-dependent nitrogenase, *Geobiol. 9*, 221–232.

26. Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2007) Clustal W and Clustal X version 2.0, *Bioinformatics 23*, 2947-2948.

27. Huelsenbeck, J. P., and Ronquist, R. (2001) MRBAYES: Bayesian inference of phylogeny, *Bioinformatics 17*, 754–755.

28. Guindon, S., and Gascuel, O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood., *Systematic Biology 52*, 696-704.

29. Abascal, F., Zardoya, R., and Posada, D. (2005) ProtTest: selection of best-fit models of protein evolution, *Bioinformatics 2*, 2104-2105.

30. Webb, C. O., Ackerly, D. D., and Kembel, S. W. (2008) Phylocom: software for the analysis of phylogenetic community structure and trait evolution, *Bioinformatics 24*, 2098-2100.

31. Nielsen, M., Lundegaard, C., Lund, O., and Petersen, T. N. (2010) CPHmodels-3.0-remote homology modeling using structure-guided sequence profiles, *Nucleic Acids Research 38*, W576-W581.

32. Georgiadis, M. M., Komiya, H., Chakrabarti, P., Woo, D., Kornuc, J. J., and Rees, D. C. (1992) Crystallographic structure of the nitrogenase iron protein from *Azotobacter vinelandii*, *Science 257*, 1653-1659.

33. Peters, J. W., Stowell, M. H., Soltis, S. M., Finnegan, M. G., Johnson, M. K., and Rees, D. C. (1997) Redox-dependent structural changes in the nitrogenase P-cluster, *Biochemistry 36*, 1181-1187.

34. Chiu, H., Peters, J. W., Lanzilotta, W. N., Ryle, M. J., Seefeldt, L. C., Howard, J. B., and Rees, D. C. (2001) MgATP-Bound and nucleotide-free structures of a nitrogenase protein complex between the Leu 127 Delta-Fe-protein and the MoFe-protein. , *Biochem. 40*, 641-650.

35. Mayer, S. M., Gormal, C. A., Smith, B. E., and Lawson, D. M. (2002) Crystallographic analysis of the MoFe protein of nitrogenase from a nifV mutant of Klebsiella pneumoniae identifies citrate as a ligand to the molybdenum of iron molybdenum cofactor (FeMoco), *J. Biol. Chem. 277*, 35263-35266.

36. Peters, J. W., Fisher, K., and Dean, D. R. (1995) Nitrogenase structure and function: a biochemical-genetic perspective, *Annu. Rev. Microbiol. 49*, 335-366.

37. Kim, J., and Rees, D. C. (1992) Crystallographic structure and functional implications of the nitrogenase molybdenum-iron protein from *Azotobacter vinelandii*, *Nature 360*, 553-560.

38. Kaiser, J. T., Hu, Y., Wiig, J. A., Rees, D. C., and Ribbe, M. W. (2011) Structure of precursor-bound NifEN: A nitrogenase FeMo cofactor maturase/insertase, *Science 331*, 91-94.

39. Ronquist, F., and Huelsenbeck, J. P. (2003) MRBAYES 3: Bayesian phylogenetic inference under mixed models, *Bioinformatics 19*, 1572-1574.

40. Huelsenbeck, J. P., and Ronquist., R. (2001) MRBAYES: Bayesian inference of phylogeny, *Bioinformatics 17*, 754-755.

41. Page, R. D. M. (1996) TREEVIEW: An application to display phylogenetic trees on personal computers, *Comp. Appl. Biosci. 12*, 357-358.

42. Xiong, J., Fischer, W. M., Inoue, K., Nakahara, M., and Bauer, C. E. (2000) Molecular evidence for the early evolution of photosynthesis, *Science 289*.

43. Sanderson, M. J. (2002) Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. , *Mol. Biol. Evol. 19*, 101-109.

44. Sanderson, M. J. (2004) r8s, version 1.70. User's manual, December 2004 ed.

45. Sanderson, M. J. (1997) A nonparametric approach to estimating divergence times in the absence of rate constancy. , *Mol. Biol. Evol 14*, 1218–1231.

46. Pagel, M. (1999) Inferring the historical patterns of biological evolution, *Nature 401*, 877-884.

47. Rodriguez-Trelles, F., Tarrio, R., and Ayala, F. J. (2002) A methodological bias toward overestimation of molecular evolutionary time scales, *Proc. Natl. Acad. Sci. USA 12*, 8112-8115.

48. Schopf, J. W. (1992) Microfossils of the early Archean Apex Chert: New evidence of the antiquity of life, *Science 260*, 640-646.

49. Godfrey, L. V., and Falkowski, P. G. (2009) The cycling and redox state of nitrogen in the Archaean ocean, *Nature Geosci 2*, 725-729.

50. Kaufman, A. J., Johnston, D. T., Farquhar, J., Masterson, A. L., Lyons, T. W., Bates, S., Anbar, A. D., Arnold, G. L., Garvin, J., and Buick, R. (2007) Late

Archean biospheric oxygenation and atmospheric evolution, *Science 317*, 1900-1903.

51. Kendall, B., Reinhard, C. T., Lyons, T. W., Kaufman, A. J., Poulton, S. W., and Anbar, A. D. (2010) Pervasive oxygenation along late Archaean ocean margins, *Nature Geosci 3*, 647-652.

52. Reinhard, C. T., Raiswell, R., Scott, C., Anbar, A. D., and Lyons, T. W. (2009) A late Archean sulfidic sea stimulated by early oxidative weathering of the continents, *Science 326*, 713-716.

53. Summons, R. E., Jahnke, L. L., Hope, J. M., and Logan, G. A. (1999) 2-Methylhopanoids as biomarkers for cyanobacterial oxygenic photosynthesis, *Nature 400*, 554–557.

54. Brocks, J. J., Buick, R., Logan, G. A., and Summons, R. E. (2003) Composition and syngeneity of molecular fossils from the 2.78 to 2.45 billion-year-old Mount Bruce Supergroup, Pilbara Craton, Western Australia, *Geochim. Cosmochim. Acta 67*, 4289–4319.

55. Eigenbrode, J. L., Freeman, K. H., and Summons, R. E. (2008) Methylhopane biomarker hydrocarbons in Hamersley Province sediments provide evidence for Neoarchean aerobiosis, *Earth and Planetary Science Letters 273*, 323-331.

56. Stams, A. (1994) Metabolic interactions between anaerobic bacteria in methanogenic environments, *Antonie van Leeuwenhoek 66*, 271-294.

57. Beiko, R. G., Harlow, T. J., and Ragan, M. A. (2005) Highways of gene sharing in prokaryotes, *Proc. Natl. Acad. Sci. USA 102*, 14332-14337.

58. Fournier, G. P., and Gogarten, J. P. (2008) Evolution of acetoclastic methanogenesis in *Methanosarcina* via horizontal gene transfer from cellulolytic *Clostridia*, *J. Bacteriol. 190*, 1124-1127.

59. Joerger, R. D., Premakumar, R., and Bishop, P. E. (1986) Tn5-induced mutants of Azotobacter vinelandii affected in nitrogen fixation under Mo-deficient and Mo-sufficient conditions, *J Bacteriol 168*, 673-682.

60. Kennedy, C., and D. Dean. 1992. The nifU, nifS and nifV gene products are required for activity of all three nitrogenases of Azotobacter vinelandii. Mol Gen Genet 231:494-8., and Dean, D. (1992) The *nifU*, *nifS* and *nifV* gene products are required for activity of all three nitrogenases of *Azotobacter vinelandii*, *Mol Gen Genet 231*, 494-498.

61. Hamilton, T. L., Ludwig, M., Dixon, R., Boyd, E. S., Dos Santos, P. C., Setubal, J. C., Bryant, D. A., Dean, D. R., and Peters, J. W. (2011) Transcriptional

profiling of nitrogen fixation in *Azotobacter vinelandii*, *J. Bacteriol. 193*, 4477-4486.

62. Joerger, R. D., and Bishop, P. E. (1988) Bacterial alternative nitrogen fixation systems *Crit. Rev. Microbiol. 16*, 1-14.

63. Burke, D. H., Hearst, J. E., and Sidow, A. (1993) Early evolution of photosynthesis: clues from nitrogenase and chlorophyll iron proteins, *Proc. Natl. Acad. Sci. USA 90*, 7134-7138.

64. Hearst, J. E., Alberti, M., and Doolittle, R. F. (1985) A putative nitrogenase reductase gene found in the nucleotide sequences from the photosynthetic gene cluster of *R. capsulata*, *Cell 40*, 219-220.

65. Staples, C. R., Lahiri, S., Raymond, J., Von Herbulis, L., Mukhophadhyay, B., and Blankenship, R. E. (2007) Expression and association of group IV nitrogenase NifD and NifH homologs in the non-nitrogen-fixing archaeon *Methanocaldococcus jannaschii*, *J Bacteriol. 189*, 7392-7398.

66. Brocker, M. J., Schomburg, S., Heinz, D. W., Jahn, D., Schubert, W. D., and Moser, J. (2010) Crystal structure of the nitrogenase-like dark operative protochlorophyllide oxidoreductase catalytic complex (ChlN/ChlB)2, *J Biol Chem 285*, 27336-27345.

67. Muraki, N., Nomata, J., Ebata, K., Mizoguchi, T., Shiba, T., Tamiaki, H., Kurisu, G., and Fujita, Y. (2010) X-ray crystal structure of the light-independent protochlorophyllide reductase, *Nature 465*, 110-114.

68. Muraki, N., Nomata, J., Ebata, K., Mizoguchi, T., Shiba, T., Tamiaki, H., Kurisu, G., . , and Fujita, Y. (2010) X-ray crystal structure of the light-independent protochlorophyllide reductase, *Nature 465*, 110-114.

69. Holland, H. D. (2006) The oxygenation of the atmosphere and oceans, *Philosophical transactions of the Royal Society of London. Series B, Biological sciences 361*, 903-915.

70. Navarro-González, R., McKay, C. P., and Mvondo, D. N. (2001) A possible nitrogen crisis for Archaean life due to reduced nitrogen fixation by lightning, *Nature 412*, 61-64.

71. Reinhard, C. T., Raiswell, R., Scott, C., Anbar, A. D., and Lyons, T. W. (2009) A late Archean sulfidic sea stimulated by early oxidative weathering of the continents, *Science 326*, 713-716.

72. Poulton, S. W., Fralick, P., and Canfield, D. E. (2010) Spatial variability in oceanic redox structure 1.8 billion years ago, *Nat. Geosci. 3*, 486-490.

73. Lovley, D. R., and Klug, M. J. (1983) Sulfate reducers can outcompete methanogens at freshwater sulfate concentrations, *Appl Environ Microbiol. 45*, 187-192.

74. Winfrey, M. R., and Zeikus, J. G. (1977) Effect of sulfate on carbon and electron flow during microbial methanogenesis in freshwater sediments., *Appl. Environ. Microbiol. 33*, 275-281.

75. Anbar, A. D. (2008) Elements and evolution, *Science 322*, 1481-1483.

76. Anbar, A. D., and Knoll, A. H. (2002) Proterozoic ocean chemistry and evolution: a bioinorganic bridge?, *Science 297*, 1137-1142.

CHAPTER 3

ENVIRONMENTAL CONSTRAINTS UNDERPIN THE DISTRIBUTION AND PHYLOGENETIC DIVERSITY OF *NIFH* IN THE YELLOWSTONE GEOTHERMAL COMPLEX

Contributions of Authors and Co-Authors

Manuscript in Chapter 3

Author: Trinity L. Hamilton

Contributions: Conceived the study, collected and analyzed output data, and wrote the manuscript.

Co-author: Eric S. Boyd

Contributions: Assisted with conceiving the study and the study design and discussed the results and implications provided valuable insight iinsight, and edited the manuscript at all stages.

Co-author: John W. Peters

Contributions: Obtained funding, discussed the results and implications, provided valuable insight, and edited the manuscript.

Manuscript Information Page

Authors: Trinity L. Hamilton, Eric S. Boyd, John W. Peters

Journal: Microbial Ecology

Status of the Manuscript:
___Prepared for submission to a peer-reviewed journal
___Officially submitted to a peer-reviewed journal
___Accepted by a peer-reviewed journal
_x_Published in a peer-reviewed journal

Publisher: Springer Science

Issue in which the manuscript appears: Volume 61, 860 – 870 (2011)

CHAPTER 3

ENVIRONMENTAL CONSTRAINTS UNDERPIN THE DISTRIBUTION AND PHYLOGENETIC DIVERSITY OF *NIFH* IN THE YELLOWSTONE GEOTHERMAL COMPLEX

Published: *Microbial Ecology*, **2011**, *61*, 860 – 870

Trinity L. Hamilton, Eric S. Boyd, John W. Peters

Department of Chemistry and Biochemistry and the Astrobiology Biogeocatalysis Research Center, Montana State University, Bozeman, Montana 59717

Abstract.

Biological nitrogen fixation is a keystone process in many ecosystems, providing bioavailable forms of fixed nitrogen (N) for members of the community. In the present study, degenerate primers targeting the nitrogenase protein-encoding gene (*nifH*) were designed and employed to investigate the physical and chemical parameters that underpin the distribution and diversity of *nifH* as a proxy for nitrogen-fixing organisms in the geothermal springs of Yellowstone National Park (YNP), Wyoming. *nifH* was detected in fifty seven of the sixty four YNP springs examined, which varied in pH from 1.90 to 9.78 and temperature from 16°C to 89°C. This suggested that the distribution of *nifH* in YNP is widespread and is not constrained by pH and temperature alone. Phylogenetic and statistical analysis of *nifH* recovered from 13 different geothermal spring environments indicated that the phylogeny exhibits evidence for both geographical and ecological structure. Model selection indicated that the phylogenetic relatedness of *nifH* assemblages could be best explained by the geographic distance between sampling sites. This suggests that *nifH* assemblages are dispersal limited with respect to the fragmented nature of the YNP geothermal spring environment. The second highest ranking explanatory variable for predicting the phylogenetic relatedness of *nifH* assemblages was spring water conductivity (a proxy for salinity), suggesting that salinity may constrain the distribution of *nifH* lineages in geographically-isolated YNP spring ecosystems. In summary, these results indicate a widespread distribution of *nifH* in YNP springs, and suggest a role for geographical and ecological factors in constraining the distribution of *nifH* lineages in the YNP geothermal complex.

Introduction

The biological reduction of dinitrogen (N_2) to ammonia (NH_3), or biological N_2 fixation (diazotrophy), is catalyzed by a diversity of microbes distributed across both the bacterial and archaeal domains. Biological nitrogen fixation is of fundamental importance in natural ecosystems, since it can relieve fixed nitrogen (N) source limitation (*1*). The molybdenum (Mo)-nitrogenase (*nif*) is the primary enzyme responsible for the conversion of N_2 to NH_3 in natural environments (*2*) and its expression is tightly regulated due to the high energetic cost (minimally 16 mol ATP are required per mol N_2 reduced) associated with this catalytic activity (*3*). In studies of diazotrophs in pure culture, N_2 fixation has been shown to be regulated from the transcriptional to post-translational levels in response to the availability of fixed sources of nitrogen (N), including ammonia (total NH_3), nitrate (NO_3^-), and nitrite (NO_2^-) (*4-8*). Thus, in environments where sources of fixed N are not limiting, a strong selective pressure would presumably exist to purge the genomes of constituent organisms of *nif*, given the high energetic cost associated with catalysis and the high metabolic cost associated with maintaining the genes necessary to synthesize an active nitrogenase (at least 12 *nif* genes need to be expressed in *Klebsiella pneumoniae*) (*9*). In contrast, in N-limited environments, a strong selective pressure would be anticipated to select for *nif* in the constituent microbial communities.

A number of studies have examined the distribution and diversity of nitrogenase in natural environments across a range of physical and chemical conditions through the use of degenerate PCR primers targeting conserved domains within the nitrogenase iron

protein (*nifH*) (*10, 11*). *nifH* in natural environments has been widely studied, with reports of high levels of diversity in the marine water column (*12*), estuarine sediments (*13*), terrestrial soils and soil rhizospheres (*14*), marine hydrothermal vents (*10*), terrestrial geothermal springs (*15*), as well as many other environments (*1*). These studies and others continue to reveal the extent of the distribution and diversity of *nifH* in natural environments; however, the physical and/or chemical parameter(s) which constrain the distribution of specific *nifH* lineages have been more difficult to determine. For example, a recent cross system analysis of *nifH* diversity from thirteen different natural environments ranging from termite hindguts to stromatolites identified patterns in the distribution of specific *nifH* lineages that only partially reflected the availability of fixed N alone, suggesting that additional unmeasured variables or a combination of measured/unmeasured variables are responsible for structuring the diversity of *nifH* in these natural systems. Indeed, a more recent investigation of *nifH* diversity in the Chesapeake Bay revealed patterns in lineage distribution that reflected the abundance of dissolved inorganic nitrogen, salinity, and dissolved organic carbon and phosphorus, indicating that a multiplicity of parameters, in addition to available fixed N, are likely to influence the phylogenetic structure of diazotrophic populations in this system (*16*).

The geothermal springs in Yellowstone National Park, Wyoming exhibit strong physical and chemical gradients, both within a given geothermal spring (*17-19*) and between geothermal springs (*20, 21*). Likewise, strong gradients of available N also exist in many of these geothermal spring environments (*20, 21*). Variability in the levels of fixed N may influence the presence or absence of diazotrophs along these gradients,

considering their potential to relieve N limitation in the ecosystem. Moreover, the physical and chemical heterogeneity in YNP geothermal spring environments would presumably create a multitude of ecological niches, some of which may be capable of supporting specific lineages of diazotrophs while limiting the growth of others (*22*). The tendency for species to retain ecological aspects of their niche over evolutionary time is termed niche conservatism (*23*), which results in discernible patterns in the distribution of phylotypes in various niches along the physicochemical gradient (e.g., ecological structure) (*24*). While patterns in the distribution of individual *nifH* lineages in natural environments have been noted previously (*1, 16*), the extent to which diazotrophic communities are ecologically structured is unknown. Here we report on the distribution and phylogenetic diversity of *nifH* along physical and chemical gradients in YNP springs. Robust phylogenetic analysis of *nifH* recovered from YNP springs coupled with community ecological tools revealed evidence for both ecological and geographic structure. Of the measured environmental variables, spring water conductivity (a proxy for salinity) was the best predictor of *nifH* phylogenetic relatedness, although only a small fraction of the variation in relatedness was explained by conductivity suggesting the importance of other unmeasured environmental parameters. In contrast a much larger fraction of the variation in *nifH* phylogenetic diversity can be attributed to the geographic distance between sites, evincing a role for dispersal limitation in structuring the *nifH* communities. In summary, the data indicate environmental parameters including salinity have imposed niche conservatism on *nifH* assemblages, and these assemblages may be dispersal limited with respect to the Yellowstone geothermal spring environment.

Methods

Sample Site Description

Four geographically-distinct sites were chosen for examining the constraints on the distribution and diversity of *nifH* in YNP springs (Table 3.1). The four sampling sites included the Witch Creek thermal field near Heart Lake (HL in the present study) in the southern part of YNP, several springs in and around Imperial Geyser located in the Lower Geyser Basin (IG) located in the central part of YNP, thermal features on the north shore of Nymph Lake (NL) located in the Northern part of the Park, and a variety of thermal features in the Back Basin of Norris Geyser Basin (NG) located in the Northern part of the Park. The thermal features sampled at HL and IG were generally neutral to alkaline in pH whereas those collected from the NG and the geothermal field located on the north shore of NL were generally acidic (Table 3.1, Figure 3.1).

Sample Collection and DNA Extraction

Samples of microbial mat and sediment were collected from a variety of locations within YNP during August of 2007. Samples that were collected within temperature and pH realms that permitted photosynthesis (*17, 25*) were of the traditional microbial mat consistency (*26*), whereas samples collected at temperature and pH combinations that did not permit photosynthesis (*17, 25*), generally consisted of sediments or iron- or sulfur-mats/filaments (*27*). Mat and sediment samples were collected aseptically using flame-

sterilized spatulas and were immediately placed in sterile microfuge tubes prior to flash freezing in a dry ice/ethanol slurry. Genomic DNA was extracted from ~100 mg of microbial mat or sediment using a combined chemical (sodium dodecyl sulfate and tris-buffered phenol) and physical (ballistic bead beating) DNA extraction protocol as previously described (*28*). Genomic DNA was quantified by agarose gel electrophoresis using the High DNA Mass Ladder (Invitrogen, Carlsbad, CA) for use in PCR. In addition to quantification, all genomic DNA extracts were subject to PCR amplification of 16S rRNA genes to ensure that extracted DNA was PCR-amplifiable. 16S rDNA amplification was performed as previously described (*29*) using ~10 ng of DNA as template and archaeal primer 931F (5'-AAGGAATTGGCGGGGGAGCA-3') or bacterial primer 1070F (5'-ATGGCTGTCGTCAGCT-3') in combination with the universal primer 1492R (5'-GGTTACCTTGTTACGACTT-3'). Both reactions were performed at an annealing temperature of 55°C.

nifH Primer Design Design and PCR Amplification

Putative *nifH* gene sequences were compiled from the GenBank database using tBLASTn with the *Nostoc* str. PCC7120 NifH sequence as bait in November of 2007. The corresponding putative *nifH* sequences were imported into MEGA4 (ver. 4.0.1) (*30*), translated in frame, and the inferred amino acid sequences were aligned using the ClustalW application (Gonnet substitution matrix, default parameters) within the MEGA4 program. Putative NifH sequences were screened for the presence of conserved signature catalytic residues including the phosphate binding P loop, the Switch regions I and II, and the Cys residues serving as ligands for the [4Fe-4S] cluster (*31, 32*). Sequences lacking

these domains were discarded without further consideration, resulting in 116 sequences. Consensus regions suitable for primer design were identified in deduced amino acid sequence alignments and reverse translation of aligned deduced amino acid sequences facilitated the design of *nifH* PCR primers containing minimal degeneracy.

Degenerate PCR primers corresponding to amino acid positions 40-49 (VGCDPKADS) and amino acid positions 157-166 (GEMMALYAAN) in the *Nostoc* str. PCC 7120 NifH sequence (NP_485497) were selected for PCR primer design. Primer specificity and PCR condition optimization was performed in reactions containing genomic DNA from positive control cultures *Azotobacter vinelandii*, *Roseiflexus* sp. Rs-1, *Synechoccocus* sp. JA-3-3Ab, and *Methanobacterium thermoautotrophicum* str. Delta H as positive amplification controls and using genomic DNA from *Thermotoga maritima* MSB8 as a negative amplification control. Primers *nifH*-119F (5'-THGTHGGYTGYGAYCCNAARGCNGAYTC-3'; 4608-fold degeneracy) and *nifH*-471R (5'- GGHGARATGATGGCNMTSTAYGCNGCNAA-3'; 3072-fold degeneracy) (Integrated DNA Technologies, Coralville, IA) were used to amplify a ~350 bp fragment of *nifH*. Fifty µL PCRs contained 10 ng of genomic DNA as template, 2 mM $MgCl_2$ (Invitrogen), 200 µM of each deoxyribonucleotide triphosphate (Promega, Madison WI), 1 µM of both the forward and reverse primer, 200 nM molecular-grade bovine serum albumin (Roche, Indianapolis, IN), and 0.25 U Taq DNA Polymerase (Invitrogen) in 1X PCR buffer (Invitrogen).

Step-down PCR conditions included an initial 4 minute denaturation at 94ºC, followed by 4 cycles of an initial denaturation at 94ºC (1 min), annealing at 68ºC, 66.5ºC, 65ºC, and 63.5ºC (1 min), and primer extension at 72ºC (1.5 min). The conditions of the

final 30 cycles consisted of denaturation at 94ºC (1 min), annealing at 60.9ºC (1 min), and primer extension at 72ºC (1.5 min), with a final extension step at 72ºC (20 min). PCR products were purified with the Wizard PCR Preps DNA Purification System (Promega, Madison, WI), quantified by agarose gel electrophoresis using the Low DNA Mass Ladder (Invitrogen), cloned using the pGEM-T Easy vector system (Promega), and sequenced using the M13F-M13R primer pair as previously described (*29*). Putative *nifH* sequences were translated using the ExPASy translate tool (http://www.expasy.ch/tools/dna.html) and the resulting translated NifH sequences were queried against the GenBank database using tBLASTn. In addition, inferred NifH amino acid sequences were aligned in Mega and evaluated for the presence of conserved catalytic signatures of *nifH*.

Chemical Analyses

Unless otherwise stated, all measurements were performed on site at the time of sample collection. The temperature and pH at each sampling site was determined using a model 59002-00 Cole-Parmer field thermometer and pH meter. An alcohol thermometer was used to confirm temperature. Initially, 64 sampling sites were chosen to sample the range of temperature and pH commonly observed in the thermal features of YNP. Thirteen of the 57 sites that yielded *nifH* PCR products (*nifH*$^{+}$) of the predicted size (~350 bp) were selected for further sequence analysis (Table 1). These sites were chosen to sample the geochemical and geographic diversity of the springs sampled in the present data set. Total sulfide (S^{2-}) and ferrous iron (Fe^{2+}) concentrations were determined on site using a Hach DR/2000 spectrophotometer (Hach Company, Loveland, CO) and the

Methylene Blue Method using Hach sulfide reagents 1 and 2 and ferrozine pillows, respectively. Samples for total ammonia, nitrate (NO_3^-), nitrite (NO_2^-), and sulfate (SO_4^{2-}) were filtered (0.22 µm) into sterile 50 mL Falcon tubes, frozen on site, and stored at -20°C prior to analysis. Concentrations of these analytes were determined colorimetrically using a SEALQuAAtro instrument (West Sussex, England) calibrated daily with fresh standards. Conductivity as a proxy for salinity was determined using a YSI model 33 S-C-T meter (Yellow Springs Instrument Company, Inc., Yellow Springs, Ohio). Conductivity values were standardized to a common temperature of 25.0°C ($Cond_{std}$) according to the formula: $Cond_{std} = Cond_T/(1-\alpha(T-25))$, where $Cond_T$ is the specific conductivity measured at ambient temperature (T) in degrees centigrade and α is the temperature correction slope factor for freshwater of 0.02. All statistical analyses and reports of conductivity repoorted herein have been standardized. Pearson correlation analyses were performed using the base package within R (version 2.10.1) (http://www.r-project.org/contributors.html). Values that were below detection were given a value of 0 in all statistical analyses.

Phylogenetic Analyses

ClustalX (ver. 2.0.9) (*33*) was used to align nucleic acid sequences using the IUB substitution matrix and default gap extension and opening penalties. A pairwise sequence identity matrix was also created using ClustalX and this was used in the program DOTUR to identify and group operational taxonomic units (OTU) using a sequence identity threshold of 99% or 0.01 (Table 3.2). Rarefaction analyses were also performed with DOTUR. The phylogenetic position of all unique *nifH* OTUs (a total of 66

sequences) identified using DOTUR (*34*) as described above was evaluated using MrBayes (version 3.1) (*35*). Tree topologies were sampled every 1000 generations for 3.75×10^6 generations using the Hasegawa, Kishino and Yano (HKY) evolutionary model with gamma-shaped rate variation with a proportion of invariable sites as recommended by Modeltest (ver. 3.8) (*36*). The protochlorophyllide reductase iron-sulfur ATP-binding gene (*bchL*) from *Chlorobium chlorochromatii* CaD3 (CP000108) and *Jannaschia* sp. CCS1 (CP000264), a related homolog of *nifH*, served as the outgroup in the phylogenetic analysis. A consensus phylogeny was generated from 782 trees sampled at stationarity (standard deviation between split frequencies <0.08). The *nifH* phylogram was rate-smoothed using the non-parametric rate smoothing (NPRS) approach implemented in the Ape (ver. 2.5) (*36*) package within R (ver. 2.10.1) and a molecular clock approach using the multidimensional version of Rambauts parameterization as implemented in PAUP (ver. 4.0) (*37*).

Community Ecological Analyses

Both NPRS and molecular clock cladograms were used to calculate Rao's quadratic entropy (D_p), the net relatedness index (NRI), and the nearest taxon index (NTI) (*38*) with Phylocom (ver. 4.0.1) (*39*). We tested whether these values differed significantly from that of a randomly assembled community using a two-tailed significance test based on 1000 independent permutations. When >975 permutations supported the observed values rather than the random or null model ($p < 0.05$), the observed rank was assumed to be significant. The results generated from the NPRS and

molecular clock cladograms were congruent (data not shown). Thus, D_p, NRI, and NTI values were averaged for both analyses and are presented in Table 3.2.

Phylocom was also used to construct a community phylogenetic distance matrix using Rao phylogenetic distances derived from both the NRPS and molecular clock *nifH* cladograms as previously described (*24*). Euclidean distance matrices derived from the nine environmental variables listed in Table 3.1 as well as a geographic distance (GD) matrix were constructed using the base package within R (version 2.10.1). A pairwise between sampling site distance matrix, as measured in km, is presented in Supp. Table 1. Environmental parameters that were below the analytical detection limit were given a value of 0 for the purpose of this study. With Rao phylogenetic distance as the response variable, model selection through Akaike Information Criteria adjusted for small sample size (AICc) and Mantel regressions were performed using the R packages Ecodist (*40*) and pgirmess (version 1.4.3) (http://pagesperso-orange.fr/giraudoux/).

Principle coordinate analysis of *nifH* assemblages was performed using the R packages vegan (http://vegan.r-forge.r-project.org/) and labdsv (http://ecology.msu.montana.edu/labdsv/R). Hierarchical cluster analysis with multiple (1000) bootstrap was performed using the R package pvclust (http://www.is.titech.ac.jp/~shimo/prog/pvclust/). Hierarchical clustering was based on Ward's agglomerative correlation method (*41*). The distance matrix described the Rao phylogenetic similarity between *nifH* communities as determined using the molecular clock smoothed cladogram served as the input for both multivariate analyses.

Nucleotide Sequence Accession Numbers

nifH nucleotide sequences from representatives of each of the 66 unique *nifH* OTUs, as identified by DOTUR at a 99% sequence identity threshold, have been deposited in the GenBank, DDBJ, and EMBL databases under accession numbers GQ426235 to GQ426277 and HM113535 to HM11357 (Table A2 in Appendix A2).

Results

Distribution of *nifH* in the YNP Geothermal Complex

DNA extracted from the 64 sediment and microbial mat samples collected from YNP springs was screened for *nifH*. Amplicons of the predicted size were generated from DNA extracted from all but 7 sites. PCR amplifications using DNA from these 7 sites were positive for 16S rRNA gene amplicons. This indicates that the DNA extracted from these sites was suitable for PCR and that *nifH* was not amplified using the degenerate primer sets or was absent in the communities sampled from these sites. The 57 geothermal environments that yielded amplicons of the predicted size nearly spanned the temperature and pH range of sites sampled in the present study (16.3°C to 89.0°C and pH from 1.90 to 9.78) (Fig. 3.1), indicating that *nifH* is widespread in YNP and that pH and temperature alone do not constrain the distribution of this gene. Six of the 7 sites that did not yield *nifH* amplicons were from high temperature (>50°C) acidic (pH <5.5) thermal features sampled from the Norris Geyser Basin and the north shore of Nymph Lake. The seventh site was a basic, high temperature (pH = 9.33, 90°C) feature in the Witch Creek area of Heart Lake. Thirteen of the 57 environmental samples where *nifH* was detected were chosen for detailed sequence and chemical analysis.

Table 3.1. Physical and chemical data for YNP thermal features where *nifH* clone libraries were constructed and sequenced. GPS coordinates for the thermal features are listed in table B3 of Appendix B.

Site[a] (YNP ID)	Temp (°C)	pH	Cond. (µmhos/ cm)	Fe^{2+} (µM)	Total Ammonia (µM)	NO_3^- (µM)	NO_2^- (µM)	Total Sulfide (µM)	SO_4^{2-} (mM)
NL13 (NA)	16.3	4.9	590	5.3	2.4	0.02	0.48	1.6	0.19
IG6 (NA)	35.5	6.4	1100	BD[c]	12.2	0.22	0.31	1.3	1.17
HL11 (NA)	38.0	9.6	2650	0.1	4.1	0.52	0.20	2.2	2.13
NG16 (NBB113	39.2	3.0	2375	19.8	94.1	0.43	0.06	3.8	1.95
IG11 (NA)	52.4	8.5	2600	BD[c]	1.1	0.34	0.16	2.5	0.20
IG5 (NA)	55.0	6.7	1500	BD[c]	8.2	0.43	0.28	0.1	1.10
NL12 (NMCN N030)	58.0	6.8	3400	1.6	2.6	0.19	0.21	4.4	0.07
NG14 (NBB113	60.2	7.5	3100	BD[c]	13.8	0.95	0.34	BD[c]	0.43
NL4	63.1	2.2	3050	15.5	26.6	BD[c]	0.33	0.9	0.51
HL5 (HLFNN 164)	69.0	9.3	3800	9.3	3.9	0.45	0.86	0.4	2.62
NG24 (NRHA0 14)	71.0	3.1	3500	11.2	71.9	0.47	0.03	5.9	1.10
NG21 (NHSP10 1)	82.0	2.6	3600	15.6	207.6	0.51	0.04	16.8	3.39
NG13 (NBB13)	82.4	7.2	6000	BD[c]	18.2	0.35	0.01	1.4	0.39

[a]NL, Nymph Lake; IG, Imperial Geyser within the Lower Geyser Basin; HL, Heart Lake Geyser Basin; NG, Norris Geyser Basin

[b]A pairwise between-site distance matrix is presented in Table 1 of the supplemental material.

[c]BD, below detection, NA, not assigned. Method detection limits for Fe^{2+} = ~180 nM; NO_3^- = ~7 nM; S^{2-} = ~150 nM

These samples were chosen in order to examine *nifH* diversity as function of the varying temperature and pH combinations commonly encountered in geothermal springs in YNP. The springs from where the samples were recovered spanned a range of temperature (16.3°C to 82.4°C) and pH (2.06 to 9.57) (Table 3.1). The physical and chemical characteristics of these thirteen springs varied substantially (Table 3.1), with concentrations of measured chemical species varying by several orders of magnitude.

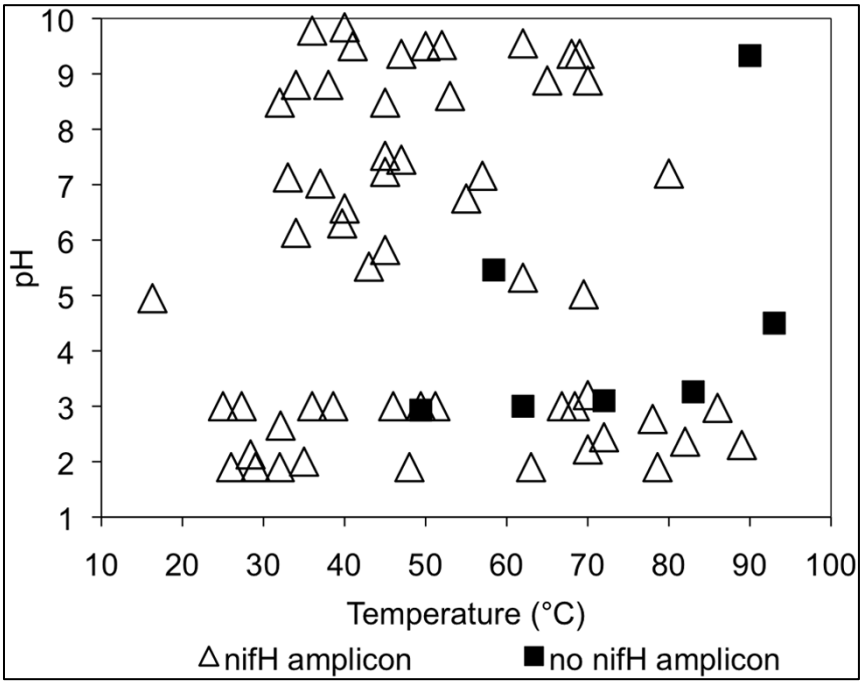


Figure 3.1. A plot of the presence/absence of *nifH* in 64 mat and sediment samples collected from 4 geographic locations in the YNP geothermal complex plotted as a function of spring water temperature and pH. Open triangles (Δ) indicate that *nifH* was detected; closed squares (■) indicate that *nifH* was not detected.

All thirteen sites contained detectable levels of total ammonia, which ranged from 1 μM to ~200 μM. Ferrous iron levels were above detection limits in eight of the thirteen sites, with the maximum measured concentration of 19.8 μM detected in an acidic spring from the Norris Geyser Basin. pH was inversely correlated with total ammonia (Pearson R^2 =

0.44, $P = 0.01$) and ferrous iron (Pearson $R^2 = 0.44$, $P = 0.01$) in the thirteen springs examined. Temperature and conductivity (a proxy for salinity) were positively correlated, (Pearson $R^2 = 0.75$, $P < 0.01$), indicative of higher conductivity with increasing temperature. Sulfate, ranging from 0.2 mM to 3.3 mM in the thirteen environments, was not significantly correlated to any of the environmental characteristics included in the analysis. None of the individual physical or chemical parameters that were measured exhibited significant autocorrelation with the geographic location of sites. In addition, the detection of *nifH* was not correlated to any individual chemical or physical parameters, including levels of fixed nitrogen, temperature or pH.

Phylogenetic Diversity of *nifH* Assemblages

nifH amplicons were sequenced from the thirteen environments chosen for detailed geochemical analysis (Table 3.1). A total of 266 *nifH* amplicons (Table 3.2, Table B2 of Appendix B) were sequenced from these geothermal communities. Phylogenetic reconstruction of the 266 *nifH* amplicons recovered from the 13 environments yielded a highly resolved and well-supported phylogram for use in examining the phylogenetic relationships of sequences from these environments.

The NRI is a measure of tree-wide phylogenetic clustering of sequences on a tree whereas the NTI metric is a standardized measure of the phylogenetic distance to the nearest taxon for each taxon within a community, which reflects the extent of terminal (branch tip) clustering (*24*). Increasingly positive NRI/NTI scores indicate that co-occurring species are more phylogenetically related than expected by chance (phylogenetic clustering).

Table 3.2. Clone library statistics for the 13 geothermal spring environments analyzed.

Geothermal Spring[a]	*n*	No. Unique Phylotypes[b]	Predicted No. Unique Phylotypes[c]	% Unique Phylotypes Sampled
NL13	14	2	2	100.0%
IG6	22	4	5	80.0%
HL11	25	10	12	83.3%
NG16	25	12	14	85.7%
IG11	21	9	12	75.0%
IG5	17	6	7	85.7%
NL12	16	7	8	87.5%
NG14	21	6	6	100.0%
NL4	18	9	11	81.8%
HL5	24	5	6	83.3%
NG24	21	8	8	100.0%
NG21	25	4	4	100.0%
NG13	17	7	8	87.5%

[a]Spring designations correspond with those presented in Table 1.
n: number of clones sequenced.
[b]Number of unique phylotypes that were sampled in the library containing *n* sequences as determined by DOTUR at a sequence identity threshold of 99%.
[c]Predicted number of unique phylotypes in the spring as determined by DOTUR at a sequence identity threshold of 99%.

In contrast, increasingly negative NRI/NTI scores indicate that co-occurring species are less phylogenetically related than expected by chance (phylogenetic overdispersion). In other words, negative NRI and NTI indicate higher than expected phylogenetic diversity in the assemblage given the species richness of that assemblage. Ten of the 13 *nifH* assemblages exhibited statistically significant and positive NRI and/or NTI values (Table 3.3). The three sites that exhibited high D_p (NG14, NG16 and NL12) exhibited NRI values that were near 0 or that were negative, suggesting that these communities are more phylogenetically overdispersed relative to the other ten assemblages.

Table 3.3. Phylogenetic diversity metrics for the thirteen *nifH* assemblages.

Plot[a]	D_p	NRI	P-*value*	NTI	P-*value*
NL13	0.14	4.38	<0.01	1.84	<0.01
HL11	0.28	3.61	<0.01	0.06	0.27
NG16	0.51	-0.28	0.28	1.13	0.09
IG6	0.06	6.82	<0.01	1.21	0.05
IG11	0.30	2.74	<0.01	0.73	0.20
NL12	0.49	0.00	0.30	1.55	0.01
IG5	0.23	3.44	<0.01	0.24	0.34
NG14	0.48	-0.12	0.32	0.82	0.17
NL4	0.30	2.75	0.01	0.09	0.45
HL5	0.13	6.18	<0.01	1.33	0.03
NG24	0.36	1.71	0.03	1.15	0.07
NG21	0.29	2.96	<0.01	1.77	<0.01
NG13	0.08	6.09	<0.01	1.55	0.01

[a]Spring designations correspond with those presented in Table 1.

All values of NTI were positive (0.06 to 1.84) (Table 3.3) indicating sequence terminal clustering of *nifH* within assemblages, as arrayed on the phylogenetic tree. Collectively, these metrics indicate that the majority of phylotypes within each YNP *nifH* assemblages are more closely related to each other than they are to the total YNP *nifH* pool, likely reflecting both long- and short-term ecological (habitat) filtering as indicated by generally positive NRI and NTI metrics, respectively.

Model selection and Mantel tests revealed that the *nifH* phylogram exhibits evidence for ecological structure, a finding that is consistent with positive NRI and/or NTI results. However, model selection and Mantel tests also indicated that the *nifH* phylogram exhibits evidence for geographic structure. The results of model selection using Rao among community distances generated from the NPRS and molecular clock

cladograms were generally congruent (data not shown); thus, ΔAICc and Mantel R^2 and associated *P*-values were averaged for both analyses and are presented in Table 3.4.

Table 3.4. Model ranking using ΔAICc and Mantel correlation coefficients (R^2) with Rao phylogenetic distance of *nifH* serving as the response variable.

Model	ΔAICc	R^2	*P*-value
GD + Cond.	0.00	0.35	<0.01
GD + Temp.	2.21	0.33	<0.01
GD	6.09	0.27	<0.01
GD + NO_2^-	7.09	0.28	<0.01
GD + Total Sulfide	7.35	0.28	<0.01
GD + NO_3^-	7.86	0.28	<0.01
GD + pH	7.92	0.28	<0.01
GD + Fe^{2+}	8.03	0.27	<0.01
GD + Total Ammonia	8.17	0.27	<0.01
GD + SO_4^{2-}	8.25	0.27	<0.01
Cond + Temp.	8.27	0.27	<0.01
Cond.	22.79	0.12	0.02
Cond. + NO_2^-	24.19	0.08	0.02
Cond. + SO_4^{2-}	24.42	0.11	0.03
Cond. + pH	24.65	0.10	0.02
Cond. + NO_3^-	24.94	0.10	0.03
Cond. + Fe^{2+}	25.07	0.10	0.03
Cond. +Total Ammonia	26.06	0.09	0.05
Temp.	26.21	0.09	0.06
Cond. + Total Sulfide	26.28	0.09	0.06
NO_2^-	26.31	0.06	0.07
NO_3^-	26.32	0.09	0.06
pH	28.33	0.03	0.12
Total Ammonia	29.12	0.02	0.18
SO_4^{2-}	29.94	0.01	0.33
Total Sulfide	30.33	0.01	0.42
Fe^{2+}	30.37	0.01	0.42

We considered the model with the lowest AICc value to be the best and evaluated the relative plausibility of each model by examining differences between the AICc value for the best model and values for every other model (ΔAICc) (*42*). Model selection

identified between-site geographic distance (GD) (ΔAICc = 6.09; Mantel R^2 = 0.27, P <0.01) as the best individual explanatory model for predicting the phylogenetic relatedness of *nifH* assemblages in YNP, although conductivity (a proxy for salinity) differences (ΔAICc = 22.79; Mantel R^2 = 0.12, P = 0.02) was also a statistically significant predictor of *nifH* phylogenetic relatedness (Table 3.4). Importantly, a model that incorporated between-site GD and conductivity differences was a more significant predictor of *nifH* phylogenetic relatedness (ΔAICc = 0.00; Mantel R^2 = 0.35, P < 0.01) than GD or conductivity alone, as indicated by a >2.0 unit increase in ΔAICc between the top ranked model and these individual models (Table 3.4). Although between-site temperature difference alone was not a statistically significant predictor of *nifH* phylogenetic relatedness (ΔAICc = 26.21, Mantel R^2 = 0.09, P = 0.06), a model that combines between-site GD and temperature differences was a statistically significant predictor of *nifH* phylogenetic relatedness (ΔAICc = 2.21, Mantel R^2 = 0.33, P < 0.01) (Table 3.4).

To further evaluate the controls on *nifH* phylogenetic diversity, principle coordinate and hierarchical cluster analysis was performed. PCO analysis indicated a strong relationship between the phylogenetic relatedness of *nifH* assemblages and the geographic proximity of the environments from where they were sampled. The first two PCO axes explained 83.7% of the variance in the phylogenetic relatedness of *nifH* assemblages between-sites (Fig. 3.2). The first axis (68.7% of variance explained) was significantly correlated with between-site GD (adj. R^2 = 0.45, P = 0.01) and to a lesser extent pH differences among sites (adj. R^2 = 0.22, P = 0.06). The second axis (14.9% of variance explained) was significantly correlated with between-site nitrite differences (adj.

R^2 = 0.38, P = 0.01). The PCO results presented here are supported by the results of agglomerative hierarchical cluster analysis (Figure B1 in Appendix B), where two well-supported clades of *nifH* assemblages were resolved. One cluster is comprised of sequences from IG and HL which are located proximal within the south central portion of YNP with the other cluster comprised of assemblages from NG and NL which are located proximal within the north central portion of YNP.

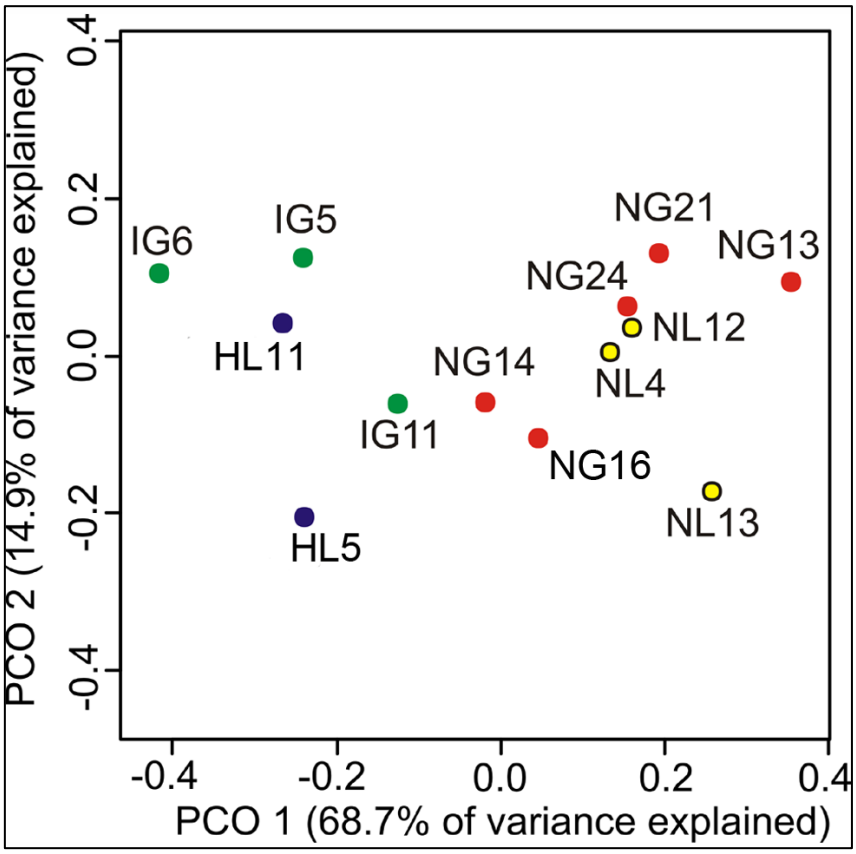


Figure 3.2. PCO analysis of *nifH* sequences from the 13 environments. The *nifH* assemblages are colored by geographic location, NG (Back Basin of Norris Geyser Basin) in red, NL (thermal features on the north shore of Nymph Lake) in yellow, IG (Imperial Geyser in the Lower Geyser Basin) in green and HL (Witch Creek thermal field near Heart Lake) in blue. Sampling site abbreviations correspond to those presented in Table 3.1.

The phylogenetic composition of *nifH* assemblages also largely reflected geographic provinces (Table B2 in Appendix B). When *nifH* sequences were classified at the order level of taxonomic resolution, and then binned based on the provinces from which they were recovered (e.g. HL, IG, NG, NL), it was clear that cyanobacteria dominated the geographically proximal HL and IG sites. In contrast, sequences affiliated with β-proteobacteria dominated the geographically proximal NG and NL sites.

Discussion

The role of chemical, physical, and/or geographical controls on microbial community structure are only beginning to be unraveled (*43, 44*). Our understanding of the influence of these parameters in structuring the diversity of functional guilds of microorganisms is even less clear (*1, 25, 43, 45*). Here, we examined the role of environmental parameters in structuring the phylogenetic diversity of the nitrogenase iron protein (*nifH*) as a proxy for diazotrophic populations inhabiting YNP geothermal environments. Given the high metabolic cost of nitrogen fixation and the number of genes required to synthesize an active nitrogenase, we hypothesized that the distribution of *nifH* would reflect the availability of sources of fixed N (NO_3^-, NO_2^-, NH_4^+) in the environments of study and that the phylogenetic structure of *nifH* genes recovered from these springs would reflect this availability over geological time scales. Surprisingly, the results presented here indicate that the potential for nitrogen fixation is widely distributed in geothermal springs across the Yellowstone geothermal complex and that this distribution is independent of the availability of N. Furthermore, the results indicate that the phylogenetic structure of diazotrophic communities reflects geographical constraints

to a greater extent than ecological constraints (e.g., fixed N availability), suggesting that the fragmented nature of Yellowstone geothermal springs imposes dispersal limitation on diazotrophic populations.

These results are consistent with a growing body of literature documenting dispersal limitation in bacterial and archaeal communities in geothermal springs in YNP. For example, a recent examination of the environmental controls on the phylogenetic diversity of the [FeFe]-hydrogenase-encoding gene as a proxy for fermentative bacteria in a number of YNP springs identified geographic distance as the best predictor of the phylogenetic diversity of HydA (*25*). Similarly, 16S rRNA gene sequences affiliated with *Sulfurihydrogenibium* (*Aquificae*) also exhibited evidence of geographic structure (*46*). Thus, bacterial ribosomal (organismal) genes and functional (trait)- encoding gene lineages in YNP springs exhibit evidence for dispersal limitation, despite the small spatial scales from which these assemblages were sampled from. In addition to the single gene analyses described above, multi-locus sequence analysis of a number of *Sulfolobus* strains from geothermal springs for the first time documented dispersal limitation in archaea inhabiting YNP springs, and suggest that the effects of dispersal limitation are observed throughout the genome of organisms (*47*). Collectively, these results indicate that the fragmented nature of geothermal provinces in YNP imparts dispersal limitation in constituent species, leading to an unknown number of endemic species in the Yellowstone geothermal complex.

It has been previously suggested that the high level of endemism in bacterial communities inhabiting YNP springs may be due to historical as well as subsequent smaller volcanic events that have together resulted in geographically-distinct provinces

(*46*) in a process analogous to that documented for land barriers such as mountains in constraining gene flow (*48*). If true, then one would expect such a phenomenon to manifest in a deep level of phylogenetic clustering in dispersal limited genes, since the last eruptive cycle creating a volcanic province (caldera) occurred roughly 600,000 years ago (*49, 50*). Indeed, analyses of the distribution of *nifH* sequences from a given assemblage on an ultrameric phylogenetic tree containing all of the *nifH* sequence assemblages (e.g., NTI and NRI values) provides strong evidence for phylogenetic structure in ten of the thirteen assemblages. Importantly, NTI values were routinely lower than NRI values in the environments sampled indicating that clustering generally occurred at a deeper level of the phylogenetic tree rather than at just the terminals, a finding that is consistent with the isolation of these lineages imposed by geographic constraints.

The phylogenetic structure of *nifH* sequence assemblages reflects, to an unknown extent, contemporary environmental constraints in addition to the historical geographical constraints as discussed above. Interestingly, the availability of fixed forms of nitrogen (NO_3^-, NO_2^-, NH_4^+) were not significant predictors of *nifH* phylogenetic relatedness, despite the high metabolic cost associated with fixing dinitrogen (minimally 16 mol ATP per mol N_2 reduced) and with maintaining the number of genes required to synthesize an active nitrogenase (at least 12 in *Klebsiella pneumoniae*) (*9*). Rather, both model selection and Mantel regressions identified spring water conductivity (a proxy for salinity) as the best predictor of the phylogenetic relatedness of *nifH* assemblages. It is important to note that salinity explained a relatively small fraction of the variance in the phylogenetic relatedness of *nifH* between sampling sites, suggesting a role for other

unmeasured biological or chemical variables in influencing the assemblages. Nevertheless, a role for salinity in constraining the phylogenetic diversity of *nifH* in YNP assemblages as determined here is consistent with the results of Moisander et al., who also identified salinity as an important driver of the phylogenetic diversity of *nifH* in Chesapeake Bay sediments (*16*).

In summary, the phylogenetic diversity of putative diazotrophs, as inferred through examination of *nifH*, was constrained at the broadest level by geographical barriers, such as the fragmented nature of geothermal provinces in the Yellowstone geothermal complex. Despite the high energetic demands associated with maintaining nitrogenase and with nitrogen fixation, our results indicate that sources of fixed nitrogen are not the primary drivers of phylogenetic diversity of putative diazotrophs in YNP. Rather, the data suggest that environmental factors including salinity impose niche conservatism on *nifH* assemblages and that this has occurred over geological timescales, perhaps in response to volcanic caldera (e.g. geographic province (*46*)) forming events. Furthermore, the data strongly suggest that diazotrophic communities are dispersal limited, perhaps due to the fragmented nature of geothermal environments within the Park.

Acknowledgments

This work was supported by the NASA Astrobiology Institute (NAI) grant NNA08C-N85A to J.W.P. T.L.H. was supported by an NSF-Integrated Graduate Educational Research and Training fellowship grant. E.S.B. graciously acknowledges support from the NAI Postdoctoral Program. We are grateful to Christie Hendrix and Stacey Gunther

for facilitating the permitting process to perform research in YNP. We are also grateful to Gill Geesey for numerous discussions and for use of laboratory facilities.

Supporting Information Available: Additional data for this study including Hierarchical clustering diagram of the relationships in D_p distance of *nifH* sequences based of on the Paup smoothed tree, pairwise distance between sampling sites in YNP, *nifH* clone library statistics and GPS coordinates for the thermal features examined. This material is available free of charge via the Internet at http://www.springerlink.com. The relevant data is located in Appendix B.

Literature Cited

1. Zehr, J. P., Jenkins, B. D., Short, S. M., and Steward, G. F. (2003) Nitrogenase gene diversity and microbial community structure: a cross-system comparison, *Environ Microbiol 5*, 539-554.

2. Bulen, W. A., and LeComte, J. R. (1966) The nitrogenase system from *Azotobacter*: two enzyme requirements for N_2 reduction, ATP dependent H_2 evolution and ATP hydrolysis, *Proc. Natl. Acad. Sci. USA 56*, 979-986.

3. Simpson, F. B., and Burris, R. H. (1984) A nitrogen pressure of 50 atmospheres does not prevent evolution of hydrogen by nitrogenase, *Science 224*, 1095-1097.

4. Daesch, G., and Mortenson, L. E. (1972) Effect of ammonia on the synthesis and function of the N 2 -fixing enzyme system in Clostridium pasteurianum, *J Bacteriol 110*, 103-109.

5. Meyer, J. (1981) Comparison of carbon monoxide, nitric oxide, and nitrite as inhibitors of the nitrogenase from Clostridium pasteurianum, *Arch Biochem Biophys 210*, 246-256.

6. Shah, V. K., Davis, L. C., and Brill, W. J. (1972) Nitrogenase. I. Repression and derepression of the iron-molybdenum and iron proteins of nitrogenase in Azotobacter vinelandii, *Biochim Biophys Acta 256*, 498-511.

7. Vaughn, S. A., and Burgess, B. K. (1989) Nitrite, a new substrate for nitrogenase, *Biochemistry 28*, 419-424.

8. Tubb, R. S., and Postgate, J. R. (1973) Control of nitrogenase synthesis in Klebsiella pneumoniae, *J Gen Microbiol 79*, 103-117.

9. Rubio, L. M., and Ludden, P. W. (2008) Biosynthesis of the iron-molybdenum cofactor of nitrogenase, *Annu Rev Microbiol 62*, 93-111.

10. Mehta, M. P., Butterfield, D. A., and Baross, J. A. (2003) Phylogenetic diversity of nitrogenase (nifH) genes in deep-sea and hydrothermal vent environments of the Juan de Fuca Ridge, *Appl Environ Microbiol 69*, 960-970.

11. Zehr, J. P., and McReynolds, L. A. (1989) Use of degenerate oligonucleotides for amplification of the nifH gene from the marine cyanobacterium Trichodesmium thiebautii, *Appl Environ Microbiol 55*, 2522-2526.

12. Langlois, R. J., LaRoche, J., and Raab, P. A. (2005) Diazotrophic diversity and distribution in the tropical and subtropical Atlantic Ocean, *Appl Environ Microbiol 71*, 7910-7919.

13. Affourtit, J., Zehr, J. P., and Paerl, H. W. (2001) Distribution of Nitrogen-Fixing Microorganisms along the Neuse River Estuary, North Carolina, *Microb Ecol 41*, 114-123.

14. Hamelin, J., Fromin, N., Tarnawski, S., Teyssier-Cuvelle, S., and Aragno, M. (2002) nifH gene diversity in the bacterial community associated with the rhizosphere of Molinia coerulea, an oligonitrophilic perennial grass, *Environ Microbiol 4*, 477-481.

15. Hall, J. R., Mitchell, K. R., Jackson-Weaver, O., Kooser, A. S., Cron, B. R., Crossey, L. J., and Takacs-Vesbach, C. D. (2008) Molecular characterization of the diversity and distribution of a thermal spring microbial community by using rRNA and metabolic genes, *Appl Environ Microbiol 74*, 4910-4922.

16. Moisander, P. H., Morrison, A. E., Ward, B. B., Jenkins, B. D., and Zehr, J. P. (2007) Spatial-temporal variability in diazotroph assemblages in Chesapeake Bay using an oligonucleotide *nifH* microarray, *Environ. Microbiol 9*, 1823–1835.

17. Havig, J. R., Raymond, J., Meyer-Dombard, D. R., Zolotova, N., and Shock, E. L. (2010) Merging Isotopes and Community Genomics in a Siliceous Sinter-Depositing Hot Spring, *Geochimica et Cosmochimica Atca*.

18. Langner, H. W., Jackson, C. R., McDermott, T. R., and Inskeep, W. P. (2001) Rapid oxidation of arsenite in a hot spring ecosystem, Yellowstone National Park, *Environ Sci Technol 35*, 3302-3309.

19. Macur, R. E., Langner, H. W., Kocar, B. D., and Inskeep, W. P. (2004) Linking geochemical processes with microbial community analysis: successional dynamics in an arsenic-rich, acid-sulphate-chloride geothermal spring, *Geobiology 2*, 163–177.

20. Nordstrom, D. K., Ball, J. W., and McClesky, R. B. (2005) Ground water to surface water: chemistry of thermal outflows in Yellowstone National Park, In *Geothermal biology and geochemistry in Yellowstone National Park* (Inskeep, W. P., and McDermott, T. R., Eds.), pp 143-162, Montana State University, Bozeman.

21. Shock E.L., Holland M., Meyer-Dombard D.R., Amend J.P., Osburn G.R., and T.P., F. (2010) Quantifying inorganic sources of geochemical energy in hydrothermal ecosystems, Yellowstone National Park, USA, *Geochimica et cosmochimica acta 74*, 4005-4043.

22. Kassen, R., and Rainey, P. B. (2004) The ecology and genetics of microbial diversity, *Annu Rev Microbiol 58*, 207-231.

23. Wiens, J. J. (2004) Speciation and ecology revisited: phylogenetic niche conservatism and the origin of species, *Evolution 58*, 193-197.

24. Webb, C. O. (2000) Exploring the Phylogenetic Structure of Ecological Communities: An Example for Rain Forest Trees, *Am Nat 156*, 145-155.

25. Boyd, E. S., Hamilton, T. L., Spear, J. R., Lavin, M., and Peters, J. W. (2010) [FeFe]-hydrogenase in Yellowstone National Park: evidence for dispersal limitation and phylogenetic niche conservatism, *ISME J Epub 2010 Jun11 doi:10.1038/ismej.2010.76*.

26. Ward, D. M., Ferris, M. J., Nold, S. C., and Bateson, M. M. (1998) A natural view of microbial biodiversity within hot spring cyanobacterial mat communities, *Microbiology and molecular biology reviews : MMBR 62*, 1353-1370.

27. Inskeep, W. P., and McDermott, T. R. (2005) Geomicrobiology of acid-sulfate-chloride springs in Yellowstone National Park, In *Geothermal biology and geochemistry in Yellowstone National Park* (Inskeep, W. P., and McDermott , T. R., Eds.), pp 143-162, Montana State University, Bozeman, MT.

28. Boyd, E. S., Jackson, R. A., Encarnacion, G., Zahn, J. A., Beard, T., Leavitt, W. D., Pi, Y., Zhang, C. L., Pearson, A., and Geesey, G. G. (2007) Isolation, characterization, and ecology of sulfur-respiring crenarchaea inhabiting acid-sulfate-chloride-containing geothermal springs in Yellowstone National Park, *Appl Environ Microbiol 73*, 6669-6677.

29. Boyd, E. S., Cummings, D. E., and Geesey, G. G. (2007) Mineralogy influences structure and diversity of bacterial communities associated with geological substrata in a pristine aquifer, *Microb Ecol 54*, 170-182.

30. Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0, *Mol Biol Evol 24*, 1596-1599.

31. Georgiadis, M. M., Komiya, H., Chakrabarti, P., Woo, D., Kornuc, J. J., and Rees, D. C. (1992) Crystallographic structure of the nitrogenase iron protein from Azotobacter vinelandii, *Science 257*, 1653-1659.

32. Walker, J. E., Saraste, M., Runswick, M. J., and Gay, N. J. (1982) Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold, *EMBO J 1*, 945-951.

33. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through

sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic Acids Res 22*, 4673-4680.

34. Schloss, P. D., and Handelsman, J. (2005) Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness, *Appl Environ Microbiol 71*, 1501-1506.

35. Huelsenbeck, J. P., and Ronquist, F. (2001) MRBAYES: Bayesian inference of phylogenetic trees, *Bioinformatics 17*, 754-755.

36. Posada, D. (2006) ModelTest Server: a web-based tool for the statistical selection of models of nucleotide substitution online, *Nucleic Acids Res 34*, W700-703.

37. Swofford, D. L. (2001) *Paup: Phylogenetic analysis using parsimony (and other methods)* Vol. 4.0b10ed., Sinauer Associate, Sunderland, MA.

38. Purcell, D., Sompong, U., Yim, L. C., Barraclough, T. G., Peerapornpisal, Y., and Pointing, S. B. (2007) The effects of temperature, pH and sulphide on the community structure of hyperthermophilic streamers in hot springs of northern Thailand, *FEMS Microbiol. Ecol. 60*, 456 - 466.

39. Webb, C. O., Ackerly, D. D., and Kembel, S. W. (2008) Phylocom: software for the analysis of phylogenetic community structure and trait evolution, *Bioinformatics 24*, 2098-2100.

40. Goslee, S. C., and Urban, D. L. (2007) The ecodist package for dissimilarity-based analysis of ecological data, *J. Stat Softw 22*, 1-19.

41. Ward, J. H. (1963) Hierarchical grouping to optimize an objective function, *Journal of American Statistical Association 22*, 1-19.

42. Johnson, J. B., and Omland, K. S. (2004) Model selection in ecology and evolution, *Trends Ecol Evol 19*, 101-108.

43. Horner-Devine, M. C., and Bohannan, B. J. M. (2006) Phylogenetic clustering and overdispersion in bacterial communities, *Ecology 87*, S100–S108.

44. Lozupone, C. A., and Rob Knight, R. (2008) Species divergence and the measurement of microbial diversity, *FEMS Microbiol Rev 32*, 557–578.

45. Boyd, E. S., Spear, J. R., and Peters, J. W. (2009) [FeFe] hydrogenase genetic diversity provides insight into molecular adaptation in a saline microbial mat community, *Appl Environ Microbiol 75*, 4620-4623.

46. Takacs-Vesbach, C., Mitchell, K., Jackson-Weaver, O., and Reysenbach, A. L. (2008) Volcanic calderas delineate biogeographic provinces among Yellowstone thermophiles, *Environ Microbiol 10*, 1681-1689.

47. Whitaker, R. J., Grogan, D. W., and Taylor, J. W. (2003) Geographic barriers isolate endemic populations of hyperthermophilic archaea, *Science 301*, 976-978.

48. Bryant, J. A., Lamanna, C., Morlon, H., Kerkhoff, A. J., Enquist, B. J., and Green, J. L. (2008) Colloquium paper: microbes on mountainsides: contrasting elevational patterns of bacterial and plant diversity, *Proc Natl Acad Sci U S A 105 Suppl 1*, 11505-11511.

49. Christiansen, R. L. (2000) Geology of Yellowstone Naitonal Park: The Quaternary and Pliocene Yellowstone Plateau Volcanic Field of Wyoming, Idaho, and Montana, (Interior, U. S. D. o. t., and Survey, U. S. G., Eds.), pp 1-146, Western Region, Menlo Park, CA, Denver, CO.

50. Lowenstern, J. B., Smith, R. B., and Hill, D. P. (2006) Monitoring super-volcanoes: geophysical and geochemical signals at Yellowstone and other large caldera systems, *Philos Transact A Math Phys Eng Sci 364*, 2055-2072.

CHAPTER 4

BIOLOGICAL NITROGEN FIXATION IN ACIDIC HIGH-TEMPERATURE GEOTHERMAL SPRINGS IN YELLOWSTONE NATIONAL PARK, WYOMING

Contributions of Authors and Co-Authors

Manuscript in Chapter 4

Author: Trinity L. Hamilton

Contributions: Conceived the study, collected and analyzed output data, and wrote the manuscript.

Co-author: Rachel K. Lange

Contributions: Collected and analyzed output data.

Co-author: Eric S. Boyd

Contributions: Obtained funding and assisted with conceiving the study and the study design, provided valuable insight, discussed the results and implications and edited the manuscript at all stages.

Co-author: John W. Peters

Contributions: Obtained funding, provided valuable insight, discussed the results and implications and edited the manuscript.

Manuscript Information Page

Authors: Trinity L. Hamilton, Rachel K. Lange, Eric S. Boyd, John W. Peters

Journal: Environmental Microbiology

Status of the Manuscript:
___Prepared for submission to a peer-reviewed journal
___Officially submitted to a peer-reviewed journal
___Accepted by a peer-reviewed journal
_x_Published in a peer-reviewed journal

Published by John Wiley and Sons, Society for Applied Microbiology and Blackwell Publishing Ltd.

Issue in which the manuscript appears: Volume 13, 2204 – 2215 (2011)

CHAPTER 4

BIOLOGICAL NITROGEN FIXATION IN ACIDIC HIGH TEMPERATURE GEOTHERMAL SPRINGS IN YELLOWSTONE NATIONAL PARK, WYOMING

Published: *Environmental Microbiology*, **2011**, *13*, 2204 – 2215

Trinity L. Hamilton, Rachel K. Lange, Eric S. Boyd*, and John W. Peters*

Department of Chemistry and Biochemistry and the Astrobiology Biogeocatalysis Research Center, Montana State University, Bozeman, Montana 59717

Abstract.

The near ubiquitous distribution of *nifH* genes in sediments sampled from fourteen high temperature (48.0°C to 89.0°C) and acidic (pH 1.90 to 5.02) geothermal springs in Yellowstone National Park, YNP suggested a role for the biological reduction of dinitrogen (N_2) to ammonia (NH_3) (e.g., nitrogen fixation or diazotrophy) in these environments. *nifH* genes from these environments formed three unique phylotypes that were distantly related to acidiphilic, mesophilic diazotrophs. Acetylene reduction assays and $^{15}N_2$ tracer studies in microcosms containing sediments sampled from acidic and high temperature environments where *nifH* genes were detected confirmed the potential for biological N_2 reduction in these environments. Rates of acetylene reduction by sediment-associated populations were positively correlated with the concentration of NH_4^+, suggesting a potential relationship between NH_4^+ consumption and N_2 fixation activity. Amendment of microcosms with NH_4^+ resulted in increased lag times in acetylene reduction assays. Manipulation of incubation temperature and pH in acetylene reduction assays indicated that diazotrophic populations are specifically adapted to local conditions. Incubation of sediments in the presence of a N_2 headspace yielded a highly enriched culture containing a single *nifH* phylotype. This phylotype was detected in all fourteen geothermal spring sediments examined and its abundance ranged from ~780 to ~6800 copies gram dry weight sediment^{-1}, suggesting that this organism may contribute N to the ecosystems. Collectively, these results for the first time demonstrate thermoacidiphilic N_2 fixation in the natural environment and extend the upper temperature for biological N_2 fixation in terrestrial systems.

Introduction

All life requires fixed sources of nitrogen (N) and its availability is what often limits productivity (*1*). Most N on Earth is in the kinetically stable form of dinitrogen (N_2), which is not bio-available (*2*). The primary enzyme that catalyzes the reduction of N_2 to bio-available NH_3 is the molybdenum (Mo)-dependent nitrogenase (Nif) although other phylogenetically-related (alternative) forms of nitrogenase that differ in their active site metal composition also likely contribute NH_3 in environments that are limiting in Mo (*3, 4*). Nif is distributed across a diversity of bacteria and the methanogenic archaea, whereas a limited subset of organisms that encode for *nif* also encode for vanadium nitrogenase or an iron-only nitrogenase (*5, 6*). Nif-catalyzed N_2 reduction is both energetically and metabolically costly, requiring a suite of gene products to assemble an active nitrogenase (at least 12 *nif* genes in *Klebsiella pneumoniae*) (*7*) and 16 moles of ATP to reduce a single molecule of N_2 (*8*). Thus, the activity of nitrogenase is tightly regulated in response to the presence of sources of fixed nitrogen, such as NH_3, NO_3^-, and NO_2^- (*9-11*).

Although essential in ecosystems where sources of fixed nitrogen are limiting, the large metabolic cost of assembling an active nitrogenase (*7*), the energetic cost of reducing N_2 (*12*), and the extreme oxygen-sensitivity of the enzyme (*13*) would seem to limit the environments where diazotrophy would occur. However, the distribution of nitrogenase genes (e.g., *nifH*) appears to be widespread in nature where they have been detected in diverse environments ranging from soils (*14, 15*) and plant nodules (*16, 17*) to the open ocean (*18, 19*) and marine hydrothermal vents chimneys (*20*). Surprisingly, a

recent examination of springs in Yellowstone National Park, WY indicated the presence of a diversity of *nifH* genes in acidic high temperature environments (pH < 3.0, Temp > 80°C) (*21*); however, this study did not address whether the organisms harboring these nitrogenase genes are active in the environment. Hyperthermophilic organisms capable of fixing N_2 at temperatures up to 92°C at circumneutral pH have been reported (*22*) and acidiphilic organisms capable of fixing N_2 at pH ~0.7 at mesophilic conditions have been described (*23*); however, thermoacidiphilic organisms capable of N_2 fixation have yet to be reported.

Here, we report the diversity, abundance, and activity of N_2 fixing organisms in sediment-associated microbial communities sampled from a number of acidic and high temperature environments in YNP. The results indicate a widespread distribution of phylogenetically-unique diazotrophic organisms that are capable of fixing N_2 under environmentally-relevant conditions. In addition, the successful enrichment of a N_2 fixating co-culture permitted an investigation into the ecological relevance of one of these novel phylotypes. Collectively, these results expand our understanding of the environmental conditions that permit biological N_2 fixation and provide the first evidence for N_2 fixation by thermoacidophilic organisms.

Materials and Methods

Sample Collection and DNA Extraction

Sediment samples were collected from thermal features in the One Hundred Springs Plain area of Norris Geyser Basin (NG) and from the thermal features along the north shore Nymph Lake (NL) in July of 2008 for use in determining the distribution,

diversity, and abundance of *nifH* genes. Sediment samples were collected aseptically with flame-sterilized spatulas, placed in sterile 1.5 mL centrifuge tubes, and flash frozen in the field using a dry ice/ethanol slurry. Samples were stored at -80°C until further processed. Genomic DNA was extracted from ~100 mg of microbial mat or sediment in duplicate using a combined chemical (sodium dodecyl sulfate and tris-buffered phenol) and physical (ballistic bead beating) DNA extraction protocol as previously described (*24*). Duplicate extracts were pooled and genomic DNA was quantified by agarose gel electrophoresis using the High DNA Mass Ladder (Invitrogen, Carlsbad, CA). To ensure that extracted DNA was suitable for PCR, all genomic DNA extracts were subjected to PCR amplification of SSU rRNA genes. SSU rDNA amplification was performed as previously described (*25*) using ~5 ng of DNA as template and archaeal primer 931F (5'-AAGGAATTGGCGGGGGAGCA-3') or bacterial primer 1070F (5'-ATGGCTGTCGTCAGCT-3') in combination with the universal primer 1492R (5'-GGTTACCTTGTTACGACTT-3'). Both reactions were performed at an annealing temperature of 55°C.

nifH Amplification and Clone Library Construction

DNA extracts that yielded a PCR product with SSU rRNA gene primers were screened for the presence of *nifH.* A step-down PCR approach was used to amplify *nifH* as previously described (*21*). The degenerate primer pair *nifH*-119F (5'-THGTHGGYTGYGAYCCNAARGCNGAYTC-3' where H=A, C, or T; Y=C or T; N=A, C, G, or T; R=A or G) and *nifH*-471R and (5'-GGHGARATGATGGCNMTSTAYGCNGCNAA-3' where H=A, C, or T; R=A or G;

N=A, C, G, or T; M=A or C; S=C or G), designed to amplify the known diversity of *nifH*, were used in PCRs containing ~10 ng of DNA with amplification and reaction conditions as previously described (*21*). PCR products were purified with the Wizard PCR Preps DNA Purification System (Promega, Madison, WI), quantified by agarose gel electrophoresis using the Low DNA Mass Ladder (Invitrogen), cloned using the pGEM-T Easy vector system (Promega), and sequenced with the M13F-M13R primer pair as previously described (*24*). Nucleotide sequences were translated in MEGA (ver. 5.0) and the inferred amino acid sequences were aligned with the ClustalW application (*26*) within MEGA (*27*). Pairwise sequence identity matrices were generated using ClustalX (ver. 2.0.9) (*26*) and these matrices were used to assign and cluster operational taxonomic units (OTUs) using DOTUR with the nearest neighbor algorithm, with a precision of 0.01 (*26*). DOTUR was also used to estimate Chao1 richness and Simpson's index of diversity. Simpson diversity indices of 1.0 indicate infinite diversity, and values of 0.0 indicate no diversity.

Chemical Analyses

Unless otherwise stated, all measurements were performed on site at the time of sample collection during July of 2008. The temperature and pH at each sampling site was determined using a model 59002-00 Cole-Parmer field thermometer and pH meter. An alcohol thermometer was used to confirm temperature. Total sulfide (S^{2-}) and ferrous iron (Fe^{2+}) concentrations were determined on site using a Hach DR/2000 spectrophotometer (Hach Company, Loveland, CO) and Hach sulfide reagents 1 and 2 and ferrozine pillows, respectively. Samples for ammonium (NH_4^+), nitrate (NO_3^-), nitrite (NO_2^-), and sulfate

(SO_4^{2-}) were filtered (0.22 µm) into sterile tubes, frozen on site, and stored at -20°C prior to analysis. Concentrations of these analytes were determined using a SEALQuAAtro instrument (West Sussex, England) calibrated daily with fresh standards. A YSI model 33 S-C-T meter (Yellow Springs Instrument Company, Inc., Yellow Springs, Ohio) was used to measure levels of conductivity.

Nitrogenase Activity Assays

The acetylene reduction assay (*28, 29*) was used as an initial proxy to assess rates of N_2 fixation in microbial communities associated with sediments collected from several acidic thermal features where *nifH* genes were detected. As an additional criterion, thermal features for acetylene reduction assays were selected in order to sample a wide range of physical and chemical conditions available in the acidic features at NG and NL. Approximately 500 mg of sediment was collected aseptically with flame-sterilized spatulas and transferred to pre-sterilized 20-mL serum vials. The sediments were immediately overlaid with 10 mL of fresh spring water sampled from where the sediments were collected. The sediment and spring water slurry were briefly purged with sterile argon gas and the vials were sealed using butyl rubber septa. When not being manipulated, serum bottles and their contents were kept in the spring source water in order to keep the bottles and their contents at ambient temperatures. Assays were initiated by injecting one mL O_2-free high purity acetylene into the headspace, yielding a headspace with final gas phase concentration of ~10% acetylene/90% argon. All assays were performed in triplicate. Killed controls were prepared in the field by syringe injection of $HgCl_2$ (1 mM final concentration). Our preliminary experiments indicated

that this concentration of $HgCl_2$ was sufficient to inhibit acetylene reduction activity, and did not significantly affect acetylene or ethylene concentrations in the assays. NH_4^+ amended assays were also prepared in the field by syringe injection of sterile O_2-free ammonium chloride (NH_4Cl) to a final concentration of 10 mM. Serum bottles were transported to Montana State University (2 hr drive) in incubated coolers in the presence of spring water such that the temperature did not change by more than 5°C during transport. At the laboratory, the assays were transferred to incubators where they were kept for the duration of sampling. The influence of spring water pH on acetylene reduction activity was assessed by adjusting the pH of the spring water using either 1M O_2-free HCl or 1M O_2-free NaOH.

The presence of ethylene in assays was determined in 50 µL subsamples of headspace gas sampled with a gas-tight syringe and measured using a model GC-8A gas chromatograph (Shimadzu Corporation, Kyoto, Japan) equipped with a 80/100 Porapak Q (Supelco, St. Louis, MO) column and flame ionization detector using helium as the carrier gas. The temperature of the injector was held at 190°C and the column temperature was held at 120°C. The concentration of ethylene was determined by comparison to freshly prepared standards. The reported rates of acetylene reduction (e.g, ethylene production) reflect the difference in ethylene concentrations between biological and killed controls as a function of incubation time. Assays were terminated when an increase of ethylene in the headspace was not observed. The time required for assays to come to completion varied between sediment samples collected from different sites. Following completion of the assays, sediments were dried at 80°C for 24 hours. All

acetylene reduction rates and qPCR template abundances for the various sediments that were assayed are normalized to g dry weight sediment (gdws).

$^{15}N_2$ Uptake Analyses

A parallel set of sediment samples were collected aseptically in pre-sterilized 20 ml serum vials and overlaid with 10 ml of spring water as described above for use in determining the uptake of $^{15}N_2$ into biomass. The spring water was briefly purged with sterile argon and the vials were capped with sterile rubber septa while under a constant stream of argon. When sealed, the serum vials were placed in spring water in order to keep the bottles and their contents at ambient source water temperature. $^{15}N_2$ uptake experiments were initiated by gas-tight syringe addition of 1.0 mL of $^{15}N_2$ (98 atom % $^{15}N_2$ gas, Sigma-Aldrich, St. Louis, MO). Six replicate $^{15}N_2$ incorporation assays were performed for each sediment sample, three of which contained 1 mM $HgCl_2$ which then served as killed controls. Incubations were also performed in the presence of unlabeled N_2. Six replicate assays were set up as described above and the experiments were initiated by gas-tight syringe injection of unlabeled ultra high purity N_2. Assays were terminated when acetylene reduction activity ceased in a parallel set of incubations (see above). To terminate assays, the bottle contents were transferred to weighing dishes and were dried at 100°C for 24 hours. After drying, samples were sent to the Stable Isotope/Soil Biology Laboratory of the Institute of Ecology at the University of Georgia in Athens, GA (http://www.swpa.uga.edu/) where they were ground using a ball-mill to a 250 μm diameter or less and were the ground sediments were then sealed in tin capsules. The contents of the tin capsules were combusted in a micro-Dumas elemental analyzer in

order to measure total N and for conversion of biomass N to N_2 for analysis with isotope mass spectrometry. One-tailed student t-tests were used to determine if increases in the ^{15}N content (atom % ^{15}N) of biomass from incubated in the presence of $^{15}N_2$, $^{15}N_2$ and $HgCl_2$, high purity N_2, and high purity N_2 and $HgCl_2$ were statistically significant.

Partial Diazotroph Enrichment Culture

A sediment slurry sampled from an unnamed spring in NG (identified here as NG4) was used to inoculate medium designed to enrich for organisms capable of reducing N_2. Base salts medium was prepared as previously described (*24*) with the following modifications: NH_4Cl was not added and the concentration of peptone was decreased from 0.05% w/v to 0.001% w/v. Medium was purged with N_2 gas passed over heated copper shavings (210°C) for 30 minutes and was then sterilized by autoclaving. Ten mL of PS medium was dispensed into 20 mL serum bottles and the gas phase was purged with sterile N_2 gas. A matrix of electron donor and acceptor combinations was tested for their ability to support diazotrophic growth. During the gas purging period, base salts medium was amended with one of the following pre-sterilized electron acceptors: elemental sulfur (final concentration of 5 g L^{-1}), sodium sulfate (final concentration of 10 mM), or sodium thiosulfate (final concentration of 10 mM). In addition, one of the following electron donors was added to the medium: butyrate, propionate, succinate, lactate, pyruvate, acetate, citrate, malate, formate, fumarate, or benzoate to a final concentration of 10 mM or glucose, galactose, fructose, lactose, ribose, maltose, mannose, arabinose, sucrose, yeast extract, peptone, Casamino Acids, tryptone, cellulose, cellobiose, or tryptone to a final concentration of 0.1% (wt/vol).

Microaerophilic growth was also tested by replacing 10% of the headspace with air. Each of the serum bottles were sealed and brought to a temperature of 80°C prior to replacement of the headspace with high purify N_2 followed by injection of ~ 100 µL of the sediment slurry. Approximately 10% of the headspace was replaced by acetylene gas, using a gas tight syringe. The growth of diazotrophs was monitored using the acetylene reduction technique as described above. Enrichments were subjected to multiple rounds of dilution to extinction and serial transfer in order to select for the best-adapted diazotroph under the various medium conditions (described above). Enrichment progress was monitored by epifluorescence spectroscopy and denaturing gradient gel electrophoresis (DGGE). DGGE was performed on PCR-amplified SSU rDNA and *nifH* fragments generated using the primers and PCR cycling conditions as described above; however, a 40 bp GC clamp was conjugated to the 5' end of the reverse primer (Integrated DNA Technologies, Coralville, IA). A vertical gradient of 45 to 70% denaturant (archaeal 16S rDNA amplicons) and 40 to 60% denaturant (*nifH* amplicons, bacterial 16S rDNA) were used to separate amplicons using DGGE (60V, 60°C for 18h) as previously described (*24*). This approach led to a highly enriched culture that harbored a single *nifH* phylotype in base salts medium containing propionate, elemental sulfur, a N_2 gas phase, and minimal yeast extract (0.001% w/v).

Quantitative PCR (qPCR)

qPCR primers for the specific amplification of the sole *nifH* gene (designated NG4 HFS) obtained from an enrichment culture that was actively reducing acetylene to ethylene (described in detail below) were designed. Primer NG HFS *nifH*-124F (5'-

CTGCATCAAAGTGCCAGGACACC -3') and NG HFS *nifH*-432R (5'-ATCGATACCTTCGTAGGCACCTTCTTCTTC -3') (Integrated DNA Technologies, Coralville, IA) were used to amplify a ~200 bp fragment of *nifH*. Fifty µL PCRs contained 10 ng of genomic DNA as template, 2 mM $MgCl_2$ (Invitrogen), 200 µM of each deoxyribonucleotide triphosphate (Promega, Madison WI), 500 nM of the forward and reverse primers, 200 nM molecular-grade bovine serum albumin (Roche, Indianapolis, IN), and 0.25 U Taq DNA Polymerase (Invitrogen) in 1X PCR buffer (Invitrogen). PCR cycling conditions consisted of an initial denaturation (94°C for 4 min) followed by 35 cycles of denaturation (94°C for 1 min), annealing (60.6°C for 1 min), and extension (72°C for 1.5 min), followed by a 20 minute, 72°C extension step. The specificity of the NG HFS-specific *nifH* primers was determined using template DNA from *Azotobacter vinelandii, Roseiflexus* sp. Rs-1, *Synechoccocus* sp. JA-3-3Ab, and *Methanobacterium thermoautotrophicum* str. Delta H, as well as from plasmid clones from a number of different environments including those obtained from alkaline environments in YNP (*21*).

qPCRs were performed in a Rotor-GeneQ quantitative real-time PCR machine (Qiagen). Reactions were performed in optically pure 0.5 mL PCR tubes (Corbett Research, Syndey, Australia). Standard curves were generated from plasmid DNA containing *nifH* from all six environments. However, since they did not differ (data not shown), only the results obtained using plasmids from NG4 are reported. The concentration of DNA in unknowns and in plasmids was quantified fluorimetrically using a NanoDrop 1000 Spectrophotometer (ThermoFisher, Waltham, MA). The SsoFast™ EvaGreen Supermix (Bio-Rad Laboratories, Hercules, CA) was used for qPCR with a

final primer concentration of 500 nM in 20 μL reactions. All reactions were performed in triplicate using the following cycling conditions: initial denaturation (95°C for 10 min) followed by 40 cycles of denaturation (95°C for 10 sec), annealing (60.6°C for 15 sec) and extension (72°C for 20 sec). Melt curves were used to verify the specificity of the qPCR reaction.

Phylogenetic Analyses

Representative *nifH* gene sequences from each of the six libraries (see above) in addition to reference and outgroup sequences were compiled and translated in frame using MEGA (ver. 4.0.1) (*27*). Protochlorophyllide reductase gene (*bchL*) sequences from *Chlorobium chlorochromatii* CaD3 (CP000108) and *Jannaschia* sp. CCS1 (CP000264) served as outgroups. Translated sequences were aligned specifying the Gonnet substitution weight matrix and default parameters using the ClustalW application within MEGA. The aligned nucleic acid sequences obtained from reverse translation of the deduced amino acid sequences were used in phylogenetic reconstructions. The phylogenetic position of environmental and enrichment *nifH* gene fragments, in relation to reference *nifH* fragments, was evaluated using MrBayes (ver. 3.1) (*30, 31*) with the GTR substitution model with gamma-shaped rate variation and a proportion of invariable sites as specified by ModelTest (ver. 3.8) (*32*). Trees were sampled every 1000 generations for 5×10^6 generations, and a consensus phylogram was generated from 500 trees sampled at stationarity (standard deviation of split replicates = 0.035) using a parameter burnin of 2.5×10^6. The consensus phylogram was projected using FigTree (version 1.2.2) (http://tree.bio.ed.ac.uk/UH).

Results and Discussion

Our previous work which revealed the widespread and near ubiquitous distribution of *nifH* genes in acidic and high temperature springs in Yellowstone National Park, Wyoming (*21*) prompted a more detailed investigation of the diversity, abundance, and potential activity of nitrogen fixing-populations in these environments. A total of 14 acidic springs in NG and NL that ranged in temperature from 48.0°C to 89.0°C and pH from 1.90 to 5.02 (Table 4.1) were examined for genetic, physiological, and isotopic evidence indicative of biological N_2 fixation.

Table 4.1. Biological, physical, and chemical data for YNP thermal features examined in the present study.

Site[a]	Temp (°C)	pH	Cond. (µmhos/ cm)	Total Ammonia (µM)	NO_2^- (µM)	NO_3^- (µM)	Total Sulfide (µM)	Fe^{2+} (µM)	SO_4^{2-} (mM)
NL5	48.0	1.9	3150	25.8	0.30	BD[b]	0.5	26.8	0.5
NG23	49.4	3.0	3175	46.5	0.06	0.27	BD[b]	8.5	1.1
NG19	51.2	3.0	3750	11.4	0.11	0.50	8.9	0.8	1.0
NG1	58.3	2.8	2600	51.7	0.07	0.59	12.6	2.3	1.4
NL1	63.1	2.2	3050	26.6	0.33	BD[b]	0.9	15.5	0.5
NG18	68.4	3.0	3750	143.9	0.03	0.35	1.1	11.0	2.5
NL12	69.5	5.0	3400	2.6	0.21	0.19	4.4	1.6	0.1
NG2	70.0	3.2	3300	103.7	0.04	0.31	14.9	4.0	1.8
NL9	70.0	2.2	4100	24.9	0.22	BD[b]	3.8	50.8	0.5
NL2	78.0	2.7	3950	41.8	0.25	BD[b]	6.4	31.6	0.4
NG3	80.0	3.1	3500	71.9	0.03	0.47	5.9	11.2	1.1
NG4	82.0	2.5	3600	207.6	0.04	0.51	16.8	15.6	3.4
NG12	83.0	3.2	2600	286	0.03	1.21	4.4	10.7	5.2
NG22	89.0	2.3	3600	202.6	0.04	0.68	12.5	2.2	2.3

[a]Abbreviations: NL, Nymph Lake; NG, Norris Geyser Basin; NA, not assigned; BD below detection

[b]Method detection limits for NO_3^- = ~7 nM; S^{2-} = ~150 nM

Fixed N sources in the 14 springs were variable, with concentrations of NH_4^+ ranging from 2.6 to 286.0 µM, concentrations of NO_2^- ranging from 0.03 to 0.33 µM, and concentrations of NO_3^- ranging from < 7 nM (detection limit) to 1.21 µM. Thus, the abundance of fixed N sources varied among the acidic and high temperature environments selected for examination.

nifH Distribution and Diversity

nifH amplicons of the predicted size (~350 bp) were detected in all of the 14 acidic springs examined in the present study, consistent with the previously documentation of a widespread distribution of *nifH* in acidic high temperature environments in YNP (*21*). Amplicons from 6 of these 14 geothermal springs, which ranged in temperature from 58.3ºC to 82.0ºC and pH from 2.22 to 3.20, were chosen for additional sequence analysis (Table 4.2).

Table 4.2. Phylogenetic diversity metrics for the six *nifH* assemblages.

Site[a]	n[b]	Chao1[c]	Simpson[c]
NG1	18	5	0.80
NL1[d]	18	9	0.87
NG2	18	13	0.83
NL2	19	12	0.87
NG3[d]	23	8	0.89
NG4[d]	25	4	0.59

[a]Abbreviations: NL, Nymph Lake; NG, Norris Geyser Basin , sites designated in Table 4.1
[b]Number of sequences in clone library
[c]Predicted Chao1 richness and Simpson diversity
[d]Data adapted from Hamilton et. al, 2010.

The predicted NifH Chao1 richness varied between 4 and 14 unique phylotypes per site, while the Simpson diversity ranged from 0.59 to 0.89 between sites. The Simpson index of diversity varied inversely (Pearson $R^2 = 0.82$, $P = 0.01$) with the NH_4^+ concentration, which indicates a greater diversity of N_2 fixing populations with decreasing concentrations of NH_4^+ in the spring water. This finding may point to a relationship between fixed N (NH_4^+) availability and nitrogenase functional diversity in these acidic environments, if functional diversity is positively correlated to genetic diversity. Increased functional diversity (variation) enables members of a community to respond differentially to changes in their environment (*33*) and has been linked to increased ecosystem function (*34*). Since nitrogenase functions to relieve fixed N limitation (*35*), this could be an example of the availability of N selecting for increased *nif* functional diversity, which would presumably result in a more resilient diazotrophic community capable of fixing N_2 under the dynamic conditions that characterize many geothermal spring environments (*36*). The YNP *nifH* genes determined in the present study were surprisingly more closely related to *nifH* from cultivated organisms than *nifH* obtained from environmental samples, suggesting that few cultivation-independent examinations of *nifH* diversity have been conducted in comparable environments. Phylogenetically, all of the *nifH* sequences clustered within one of three lineages (Fig.4.1). *nifH* sequences within cluster 1 were recovered from all six environments, where they represented between 4.8% and 38.9% of the total clones recovered from each of the environments. This cluster of sequences was distantly related (72-79% sequence identities) to *nifH* from the α-proteobacterium *Bradyrhizobium japonicum*, an organism that characteristically inhabits root nodules (*37*).

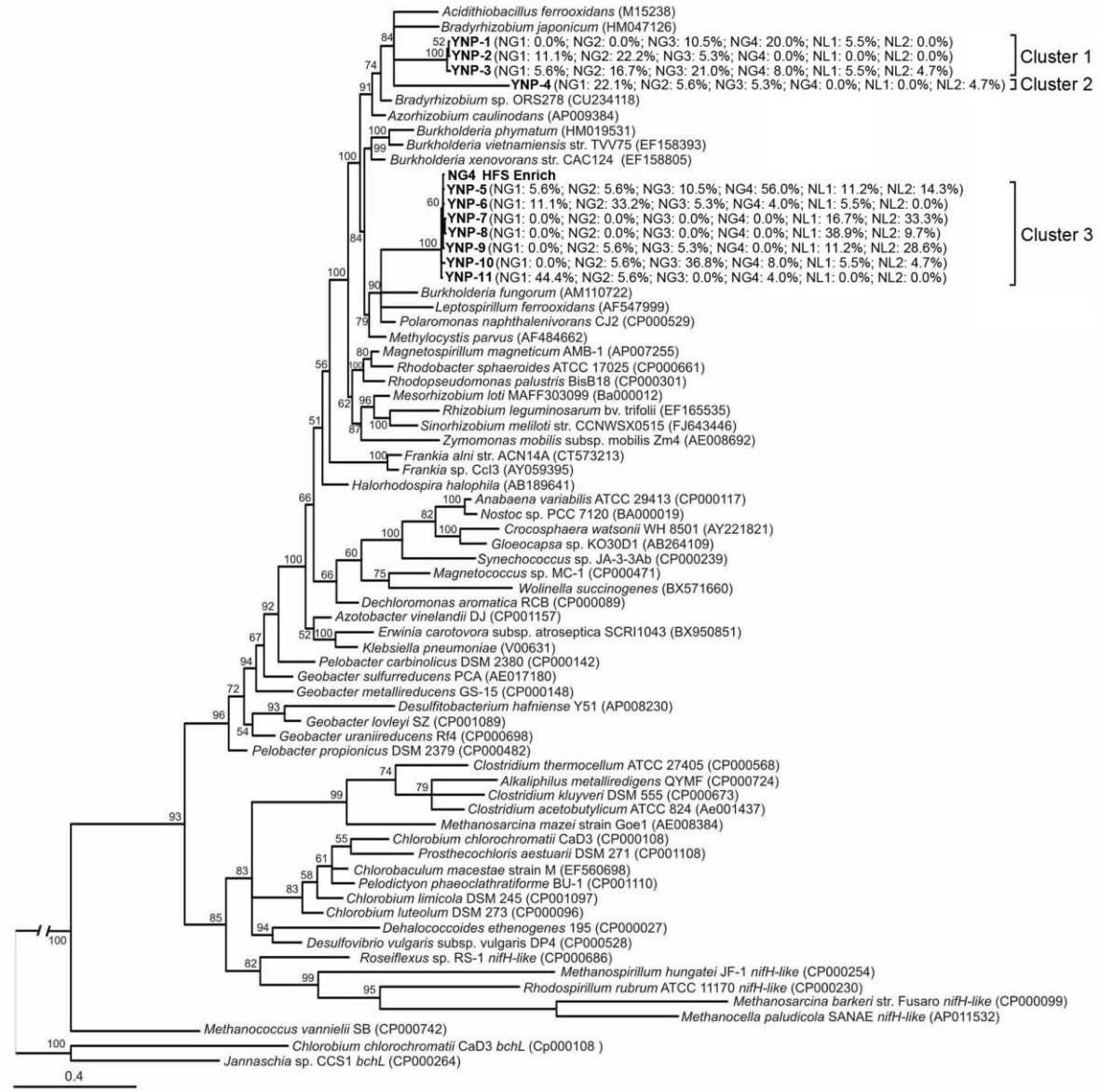


Figure 4.1. Bayesian inferred phylogenetic reconstruction of *nifH* gene fragments recovered from YNP geothermal spring environments and from YNP geothermal spring enrichments (bold-faced). Sequence abundance of each phylotype in each environment is indicated in parantheses. Sequence designations correspond with those presented in supplemental table 1. The tree is rooted with *bchL* genes. The scale bar indicates 4 substitutions per 10 positions; posterior probabilities are indicated at nodes.

The sole *nifH* sequence in cluster 2 was obtained from NG2, which was 72.3% identical to *B. japonicum*. Sequences comprising cluster 3 were also identified in all six sites, where they represented between 55.6% and 90.5% of the total clones recovered from each of the environments. These sequences were 82% identical to *nifH* from the α-proteobacterium *Polaramonas naphthalenivorans*. Importantly, despite repeated attempts, phylogenetic reconstructions could not resolve two key polytomous nodes in the phylogenetic tree with relevance to the origin and evolution of acidiphilic N_2 fixation in cultivated and environmental sequences. The first polytomy comprised *nifH* from the acidiphile *Acidithiobacillus ferroxidans*, *B. japonicum*, and YNP *nifH* clusters 1 and 2. The second polytomy comprised *nifH* from the acidiphile *Leptospirillum ferrooxidans*, *Burkholderia fungorum*, *P. napthalenivorns*, and YNP *nifH* cluster 3. Phylogenetic reconstructions with additional reference sequences and/or longer *nifH* fragments will be required to further evaluate the origin and evolution of N_2 fixation among acidiphiles.

Potential Biological N_2 Reduction Rates

The near ubiquitous occurrence of phylogenetically-unique *nifH* lineages in the six acidic high temperature YNP geothermal springs examined in the current study, and the lack of representatives in other thermal and non-thermal environments and cultivated organisms, implies that these phylotypes may have evolved in environments such as these and that the *nif* genes are likely to be expressed and active, at least transiently. This hypothesis is supported by the elevated energetic costs of maintaining the genes required to synthesize an active nitrogenase complex (at least12 *nif* in *Klebsiella pneumoniae*) (*7*) which would presumably represent a competitive disadvantage to maintain, if the

organism were never to utilize the gene products (*38*). We evaluated this hypothesis by examining the potential for biological N_2 reduction in the sediment-associated assemblages using the acetylene reduction technique which has been extensively used as a proxy for assessing N_2 reduction potentials (*28*).

Ethylene, the product of acetylene reduction, was detected in the headspace of all assays containing sediment samples from which *nifH* genes were detected (Table 4.4). The rates of acetylene reduction associated with sediments sampled from each site varied from 0.48 ± 0.01 nmol $gdws^{-1}$ day^{-1} to 1.53 nmol $gdws^{-1}$ day^{-1} in unamended assays (Table 4.3). Rates of acetylene reduction in NH_4^+ amended assays were similar to rates in unamended assays (slope of linear relationship = 0.94, adj. R^2 = 0.91), both of which were significantly correlated with NH_4^+ concentration (Pearson R^2 = 0.83 (NH_4^+ amended) and 0.74 (unamended); P = 0.01 (NH_4^+ amended) and 0.02 (unamended), respectively). Acetylene reduction rates for symbiotic diazotrophic populations and oceanic nitrogen fixation have been well-documented (*35*); however, fewer studies have reported this data for sediment-associated microcosms. The acetylene reduction rates associated with the sediments from high temperature acidic springs reported here are lower than previous reports of acetylene reduction rates in sediment-associated microbial communities. For example, acetylene reduction rates ranged from ~30 – 85 μmol C_2H_4 $gdws^{-1}$ day^{-1} in sediments sampled from salt marsh environments (*39*). However, the rates determined here were comparable to acetylene reduction rates measured from lake sediment from Lake St. George, Ontario, Canada, ~0.2 to 2.7 nmol C_2H_4 $gdws^{-1}$ day^{-1} (*40*).

Table 4.3. Acetylene reduction rates.

	Experimental Treatment		
Site[a]	Unamended (nmol gdws^{-1} day^{-1})	NH_4^+amended (nmol gdws^{-1} day^{-1})	Lag Time[a] (days)
NG1	0.62 ± 0.02	0.37 ± 0.02	4
NL1	0.71 ± 0.03	0.76 ± 0.01	2
NG2	0.82 ± 0.02	0.72 ± 0.01	3
NL2	0.59 ± 0.02	0.43 ± 0.01	2
NG3	0.48 ± 0.01	0.46 ± 0.02	3
NG4	1.53 ± 0.01	1.42 ± 0.01	2

[a]Abbreviations: NL, Nymph Lake; NG, Norris Geyser Basin , sites designated in Table 4.1
[b]Number of days before detectable ethylene in NH_4^+-amended assays, relative to unamended assays.

The primary difference observed between the amended and unamended assays was in the length of time before acetylene reduction activity could be detected. The length of the lag times in both NH_4^+ amended and unamended assays was negatively correlated with the rate of acetylene reduction [(Pearson R^2 = 0.29 (NH_4^+ amended) and 0.16 (unamended)]. Collectively, these findings may indicate a relationship between the rate of biological NH_4^+ utilization (NH_4^+ sink) in the community, and the rate of N_2 reduction (NH_3 source) by the diazotrophic component of the community.

The influence of incubation temperature and spring pH on acetylene reduction activity was examined in sediments from NG4, since these yielded the highest rates of activity (Table 4.3). Acetylene reduction activity by sediments sampled from NG4 (Fig. 4.2) was sensitive to the pH of the growth medium as well as the incubation temperature. Maximum rates of acetylene reduction (18.5 nmol C_2H_4 / µg DNA) occurred at an incubation temperature of 80°C and at pH 2.50, which are similar to the ambient

temperature (82°C) and pH (2.55) of NG4 where the sediments were collected (Table 4.1). Acetylene reduction activity appeared to be more sensitive to temperature than to pH, as indicated by an ~80% and 100% reduction when the incubation temperature was adjusted to 70°C and 90°C, respectively when compared to an ~75% and 50% decrease in activity when the pH was adjusted to 1.5 and 4.5, respectively (Fig. 4.2).

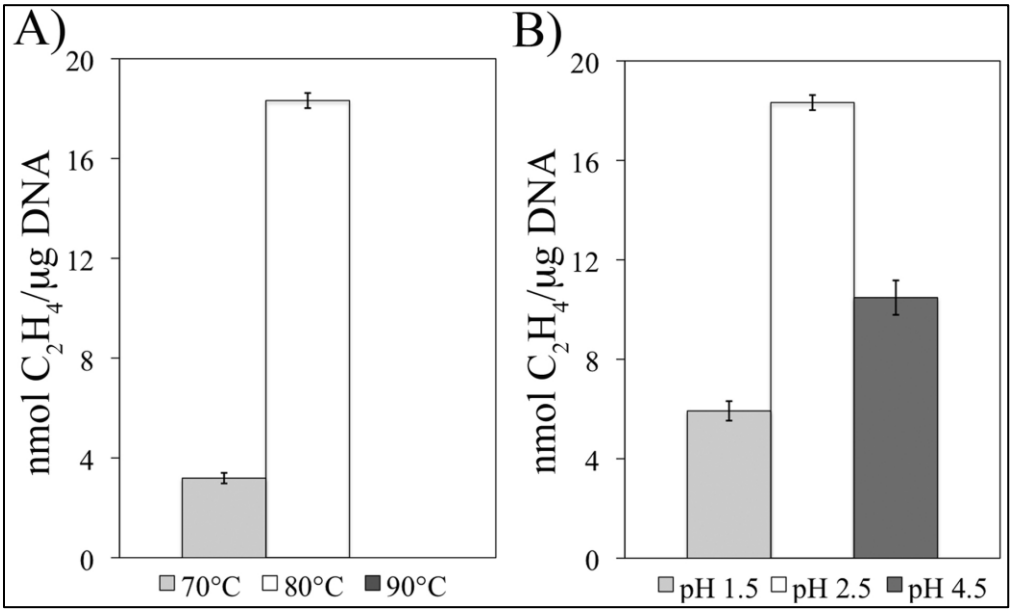


Figure 4.2**.** Acetylene reduction activity associated with sediment biomass sampled from NG4 as a function of incubation temperature and pH of medium. All treatments were run in triplicate, and the standard deviation among replicates is indicated. Acetylene reduction activity was not detected when incubated at 90°C and pH 2.5.

The observation that acetylene reduction activity did not respond linearly to increasing temperature or pH provides additional evidence that these reactions are biologically-mediated, rather than catalyzed abiotically. Moreover, these results indicate that the diazotrophic population(s) responsible for this activity is likely to be adapted to the environmental conditions from which it was sampled.

Isotopic Studies of Biological N_2 Fixation

The natural abundance of N isotopes (δ15N versus air) in biomass associated with sediments from the six environments subjected to *nif* sequence analysis were found to be isotopically-light (range: -1.493 to -2.451) (Table 4.4) which is suggestive of biomass N that is at least partially derived from biological N_2 fixation (*41*). However, it is unknown if the N is derived from allochthonous or autochthonous sources (*42*).

$^{15}N_2$ tracer experiments were used to further examine the potential for N_2-fixing activity in populations associated with sediments sampled from acidic high temperature YNP springs to contribute autochthonous fixed N to the local ecosystem. However, given previous reports of abiological N_2 fixation into organic N in experimental conditions similar to those present in the environments subject to the current study [e.g., ferruginous, sulfidic, presence of mineral sulfide solid phases, high temperature) (*43, 44*)], it was necessary to determine this potential prior to determinations of the biological potential for N_2 reduction. To achieve this, we compared the atom $\%^{15}N$ in biomass incubated in the presence of $^{15}N_2$ or high purity N_2 in assays amended with 1 mM $HgCl_2$ (killed controls). If significant abiological N_2 fixation was occurring, then the atom $\%^{15}N$ would be expected to be elevated in assays incubated in the presence of $^{15}N_2$ when compared to those incubated in the presence of high purity N_2. With the exception of NL2, the atom $\%^{15}N$ was never greater in killed control assays incubated in the presence of $^{15}N_2$ when compared to high purity N_2 (Fig. 4.3), indicating that significant abiological N_2 fixation is not occurring. Biomass sampled from NL2 exhibited elevated atom $\%^{15}N$ values in the presence of $^{15}N_2$ when compared to high purity N_2, although the increase was not statistically significant ($P = 0.08$) (Table 4.4). Further examination of abiological N_2

fixation in this and other similar environments is warranted, but is beyond the scope of this study.

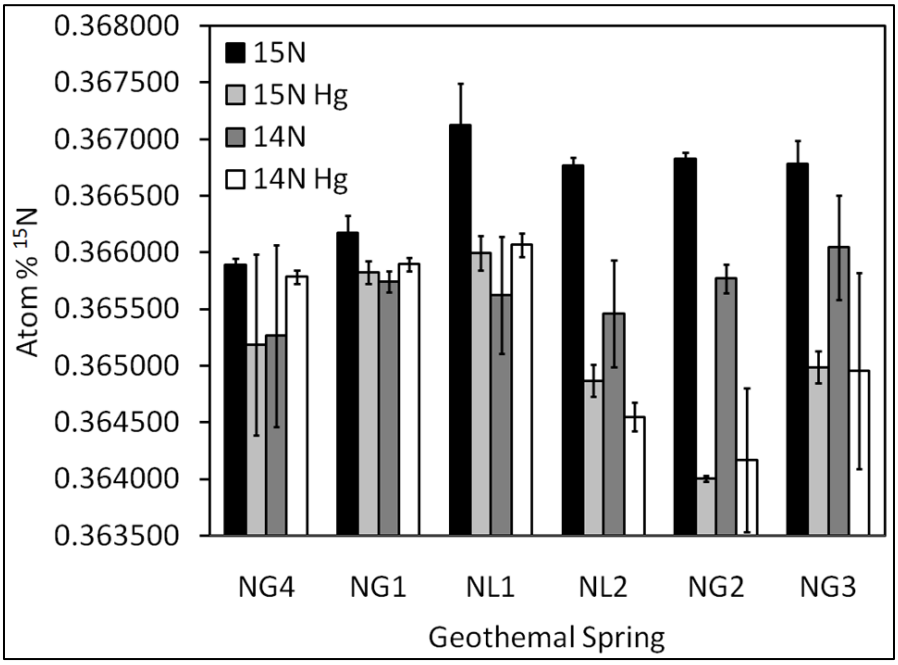


Figure 4.3. Biomass atom %15N in assays incubated in the presence of $^{15}N_2$ (15N), in the presence of $^{15}N_2$ and 1 mM $HgCl_2$ (15N Hg), in the presence of high purity N_2 (14N) and in the presence of high purity N_2 and 1 mM $HgCl_2$ (14N Hg) for six geothermal spring sediments (see Table 1 for site descriptions).

The observation that biomass atom %^{15}N was not significantly greater in killed-controls incubated in the presence of $^{15}N_2$ when compared to high purity N_2 indicates that the difference in atom %^{15}N in biomass incubated in the presence of $^{15}N_2$ from assays incubated in the absence and presence of $HgCl_2$ can be attributable to biological activity. Biomass sampled from all environments when incubated in the presence of $^{15}N_2$ was significantly enriched ($P < 0.05$) in ^{15}N when compared to biomass incubated in the presence of $^{15}N_2$ and $HgCl_2$, with the exception of NG1 ($P = 0.06$) (Fig. 4.3, Table 4.4).

This indicates that populations associated with the sediments sampled from each environment are capable of fixing N_2 into biomass. A comparison of atom $\%^{15}N$ in biomass incubated in the presence of $^{15}N_2$ and in the presence of high purity N_2 can be used to further evaluate the potential for N_2 fixation and to estimate the relative potential for biological N_2 fixation among sites, assuming that all sites contain similar amounts of biomass.

Table 4.4. $^{15}N_2$ incorporation assays[a].

	Geothermal Spring[c]					
t-test[b]	NG4	NG1	NL1	NL2	NG2	NG3
$^{15}N_2$ vs. $^{15}N_2$ Hg	0.04	0.06	0.01	0.00	0.00	0.01
$^{15}N_2$ vs. $^{14}N_2$	0.05	0.00	0.05	0.02	0.00	0.09
$^{14}N_2$ vs. $^{14}N_2$ Hg	0.09	0.04	0.12	0.03	0.03	0.02
$^{15}N_2$ Hg vs. $^{14}N_2$ Hg	0.07	0.19	0.34	0.08	0.35	0.48

	Biomass Characteristics					
Measurement	NG4	NG1	NL1	NL2	NG2	NG3
Biomass $\delta^{15}N$ versus Air	-1.882 (0.221)	-1.808 (0.407)	-2.451 (0.336)	-1.901 (0.265)	-1.493 (0.208)	-1.873 (0.040)
Biomass % total N	0.016 (0.002)	0.018 (0.003)	0.014 (0.005)	0.027 (0.004)	0.0120 (0.001)	0.010 (0.001)

[a]Values reported for 3 replicates with exception of NG4 where 6 replicates were run,
[b]*P*-values for one tailed t-tests are reported for each comparison for biomass sampled from each geothermal spring
[c]Abbreviations: NL, Nymph Lake; NG, Norris Geyser Basin , sites designated in Table 4.1

To the extent that total N abundance is a proxy for biomass, the latter assumption would be supported by biomass total N abundances associated with each site that varied by no more than a factor of three across sites (Table 4.3). Biomass from assays incubated in the presence of $^{15}N_2$ were significantly enriched ($P < 0.05$) in ^{15}N when compared to biomass

from assays incubated in the presence of high purity N_2 with the exception of NG3 (P = 0.09). Increasing enrichment in ^{15}N in sediment-associated biomass followed the trend NL1> NL2> NG2> NG3> NG4> NG1 (Fig. 4.3), which suggests different potentials for biological N_2 fixation at each site. The difference in biomass atom $\%^{15}N$ in assays incubated in the presence of $^{15}N_2$ when compared to high purity N_2 varied inversely with NO_3^- (Pearson R^2 = 0.95, P = 0.01) and positively, albeit to a lesser extent, with NO_2^- (Pearson R^2 = 0.70, P = 0.04). These results suggest that the diazotrophic component of these communities is responding to N availability, albeit differentially in response to the availability of NO_3^- and NO_2^-.

Enrichment of Thermoacidiphilic Diazotrophic sp. NG4 HFS

Sediments from NG4 (pH 2.55, 82.0°C) were used in efforts to enrich for organisms capable of N_2 reduction, since microcosms containing these sediments exhibited the highest rates of acetylene reduction activity among the six sites examined (Table 2). Sediments from NG4 were inoculated into an anaerobic base salts medium (pH 2.5) that contained propionate, elemental sulfur (S^o), and yeast extract (YE) under a 10% acetylene/90% N_2 headspace. With the exception of the organic N likely to be present in YE and the N_2 headspace, no additional N was provided to the enrichment. Incubation of the enrichment at 80°C resulted in the production of ethylene gas. Dilution to extinction in combination with systematic decreases in YE concentration in serial transfers to a final concentration of 0.001% w/v resulted in a highly enriched culture (named NG4 HFS) containing a single *nifH* phylotype (Fig. 4.1). The *nifH* phylotype grouped with YNP cluster 3 *nifH*, which were the dominant *nifH* sequences in all

libraries examined, including NG4 where the sediment was sampled. Similar to the acetylene reduction activity of sediments sampled from NG4, acetylene reduction activity by NG4 HFS was sensitive to the pH of the growth medium as well as the incubation temperature (data not shown). The culture has been maintained the lab since July 2008 on propionate, S°, and trace YE, but will not grow on other carbon sources in combination with any of the electron acceptors tested (see Materials and Methods).

The current upper temperature limit for biological N_2 fixation is 92°C, and this activity was documented in a methanogenic archaean isolated from thermal effluent from a deep sea hydrothermal vent (*22*). The upper temperature limit for N_2 fixation activity (measured using the acetylene reduction technique) among bacteria is 63.4°C, which was documented in strains of the cyanobacterium *Synechococcus* in an alkaline mat community in YNP (*29*). Thus, the acetylene reduction activity documented for the NG4 HFS enrichment culture at 80°C and the acetylene reduction activity and $^{15}N_2$ incorporation into biomass at 82°C as demonstrated herein represents a new upper temperature limit for bacterial nitrogen fixation.

In addition, the only organisms capable of N_2 fixation at acidic pH are *Acidithiobacillus ferrooxidans* (*45*), *Leptospirillum ferrooxidans* (*45*), and *Leptospirillum ferrodiazotrophum* (*23*); none of which are thermophilic. *nifH* sequences from *Acidithiobacillus ferroxidans* and *Leptospirillum ferrooxidans* were among the most closely related, albeit at ~80% sequence identities, to clusters 1, 2, and 3 *nifH* sequences recovered from YNP springs (Fig. 4.1). Thus, the acetylene reduction activity documented for the NG4 HFS enrichment culture at pH 2.50 and the acetylene reduction

activity and $^{15}N_2$ incorporation into biomass at a pH of 2.55 in the present study indicates that acidiphilic N_2 reduction can occur at high temperature. To the authors knowledge, this is the first documentation of a thermoacidiphilic N_2 fixation.

Ecological Relevance of NG4 HFS to Acidic High Temperature YNP Spring Communities

nifH from enrichment NG4 HFS formed a cluster with environmental clones obtained from all six environments where *nifH* genes were sequenced, indicating the potential widespread distribution of this phylotype in acidic and high temperature springs in YNP.

Table 4.5. Abundance of phylotype NG4 HFS *nifH* in YNP geothermal spring environments as assessed using qPCR.

Site[a]	Temp (°C)	pH	Copies[b] $gdws^{-1}$
NL5	48.0	1.90	784 ± 14
NG23	49.4	3.00	1490 ± 68
NG19	51.2	3.00	1605 ± 23
NG1	58.3	2.81	2093 ±16
NL1	63.1	2.22	2089 ± 50
NG18	68.4	3.00	3370 ± 28
NL12	69.5	5.02	891 ± 35
NG2	70.0	3.20	3943 ± 62
NL2	78.0	2.77	3791 ± 70
NL9	70.0	2.22	4130 ± 25
NG3	80.0	3.12	4650 ± 46
NG4	82.0	2.55	4101 ± 57
NG12	83.0	3.26	1867 ± 66
NG22	89.0	2.30	6822 ± 13

[a]Abbreviations: NL, Nymph Lake; NG, Norris Geyser Basin , sites designated in Table 4.1
[b]$gdws^{-1}$: g dry weight $sediment^{-1}$

qPCR with NGB HFS-specific *nifH* primers was used to examine the distribution and abundance of this phylotype in fourteen geothermal springs in NG and NL (Table 4.1), including the six where *nifH* sequence and activity data was obtained. NGB HFS *nifH* phylotypes were detected in all 14 environments examined, which ranged in temperature from 48.0°C to 89.0°C and pH from 1.90 to 5.02. The abundance of phylotype NG4 HFS ranged from ~800 copies $gdws^{-1}$ to ~6800 copies $gdws^{-1}$ (Table 4.5). The abundance of NG4 HFS *nifH* genes varied positively and to a significant extent with spring temperature (Pearson $R^2 = 0.55$, $P < 0.01$), which suggests that NG4 HFS is better adapted to environments in NG and NL with high temperature. No other environmental parameter correlated significantly to the abundance of NG4 HFS *nifH*.

Conclusions

Three novel lineages of *nifH* genes were identified in the present study that were widely distributed in each of the six springs examined, which may suggest that these lineages are adapted to acid conditions. The positive and statistically significant correlation between NH_4^+ abundance and rate of acetylene reduction may point to a dynamic relationship between sources of N (N_2 fixation or NH_3 production) and sinks (NH_4^+ consumption) in the system, which could explain why investigations of the parameters which constrain *nifH* diversity have failed to reveal statistically significant correlations with concentrations of nitrogenous in the environment (*21, 35*). Coupled analyses of N flux, the diversity of expressed *nif* genes, and *in situ* microbial activity measurements will be required to determine the relationship between these variables in future analyses.

In the present study, the natural abundance of N isotopes (δ15N versus air) in biomass associated with sediments sampled from acidic and high temperature environments were found to be negative (range: -1.493 to -2.451) which is suggestive of biomass N that is derived from biological N_2 fixation (*41*). The δ15N in biomass collected from a wide range of thermal features in YNP indicates values that range from negative to well above zero; however, these data were collected from chemotrophic and phototrophic communities inhabiting circumneutral to alkaline environments (*42, 46*). While our experimental design cannot preclude the possibility that the isotopically-light biomass N could result from allochotonous input from organics in surface soil or particulates from aeolian deposition (*42*), our results indicate the potential for autochthonous fixed N resulting from biological N_2 fixation, which may also contribute to the isotopically-light biomass N in these systems.

Previous reports of biological N_2 fixation at temperatures up to 92°C were conducted at circumneutral pH (*22*). Likewise, demonstrations of N_2 fixation under extreme acid conditions down to pH of 0.7 were conducted at mesophilic temperatures (*23*). The demonstration of biological N_2 fixation under a combination of high temperature (82°C) and acidic pH (2.55) as reported herein is the first report of N_2 fixation under these conditions. The present upper temperature limit for N_2 fixation (92°C) was reported in an organism isolated from a marine hydrothermal vent. The upper temperature limit for N_2 fixation in terrestrial environments is 63.4°C (*29*). Thus, the activity detected in the present study represents a new upper temperature limit (82°C) for biological N_2 fixation in terrestrial environments and represents a new upper temperature

limit for N_2 fixation among bacterial diazotrophs. Importantly, the detection of ~6800 copies of *nifH* NG4 HFS $gdws^{-1}$ in sediments from a geothermal spring in YNP with a temperature of 89°C and a pH of 2.30 (Table 4.5) indicates that this activity may extend well beyond the upper temperature limit for terrestrial biological N_2 fixation of 82°C determined in the present study. Likewise, the detection of ~780 copies of *nifH* NG4 HFS $gdws^{-1}$ in sediments from a geothermal spring in YNP with a pH of 1.90 (Table 4.5) and acetylene reduction activity in an environment with a pH of 2.22 (Table 4.3) indicates that this activity may extend well into more acidic pH realms in the Yellowstone geothermal complex. Collectively, these results point to a role for biological N_2 fixation in contributing fixed N to microbial communities inhabiting acidic and high temperature environments in YNP.

Acknowledgements

This work was supported by the NASA Astrobiology Institute (NAI) grant NNA08C-N85A to J.W.P and the National Science Foundation (PIRE-0968421) to J.W.P. T.L.H. was supported by an NSF-Integrated Graduate Educational Research and Training fellowship grant. E.S.B. graciously acknowledges support from the NAI Postdoctoral Program. We are grateful to Christie Hendrix and Stacey Gunther for facilitating the permitting process to perform research in YNP. We are also grateful to Gill Geesey for numerous discussions and for use of laboratory facilities.

Supplementary information: The relevant supplemental table describing the *nif*clones recovered and the GPS coordinates of the thermal featuers examined in YNP in the present study is located in Appendix C.

Literature Cited

1. Falkowski, P. G. (1997) Evolution of the nitrogen cycle and its influence on the biological sequestration of CO2 in the ocean, *Nature 387*, 272-275.

2. Howard, J. B., and Rees, D. C. (1996) *Strucutral basis of biological nitrogen fixation, Chem Rev 96*, 2965-2982.

3. Joerger, R. D., and Bishop, P. E. (1988) Bacterial alternative nitrogen fixation systems, *Crit Rev Microbiol 16*, 1-14.

4. Kessler, P. S., McLarnan, J., and Leigh, J. A. (1997) Nitrogenase phylogeny and the molybdenum dependence of nitrogen fixation in Methanococcus maripaludis, *J Bacteriol 179*, 541-543.

5. Bishop, P. E., Jarlenski, D. M., and Hetherington, D. R. (1982) Expression of an alternative nitrogen fixation system in Azotobacter vinelandii, *J Bacteriol 150*, 1244-1251.

6. Chisnell, J. R., Premakumar, R., and Bishop, P. E. (1988) Purification of a second alternative nitrogenase from a nifHDK deletion strain of Azotobacter vinelandii, *J Bacteriol 170*, 27-33.

7. Rubio, L. M., and Ludden, P. W. (2008) Biosynthesis of the iron-molybdenum cofactor of nitrogenase, *Annu Rev Microbiol 62*, 93-111.

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

9. Buchanan-Wollaston, V., Cannon, M. C., Beynon, J. L., and Cannon, F. C. (1981) Role of the nifA gene product in the regulation of nif expression in Klebsiella pneumoniae, *Nature 294*, 776-778.

10. Dodsworth, J. A., Cady, N. C., and Leigh, J. A. (2005) 2-Oxoglutarate and the PII homologues NifI1 and NifI2 regulate nitrogenase activity in cell extracts of Methanococcus maripaludis, *Mol Microbiol 56*, 1527-1538.

11. Lowery, R. G., Saari, L. L., and Ludden, P. W. (1986) Reversible regulation of the nitrogenase iron protein from Rhodospirillum rubrum by ADP-ribosylation in vitro, *J Bacteriol 166*, 513-518.

12. Simpson, F. B., and Burris, R. H. (1984) A nitrogen pressure of 50 atmospheres does not prevent evolution of hydrogen by nitrogenase, *Science 224*, 1095-1097.

13. Gallon, J. R. (1981) The oxygen sensitivity of nitrogenase: a problem for biochemists and micro-organisms, *Trends in Biochemical Sciences 6*, 19-23.

14. Shaffer, B. T., Widmer, F., Porteous, L. A., and Seidler, R. J. (2000) Temporal and Spatial Distribution of the nifH Gene of N(2) Fixing Bacteria in Forests and Clearcuts in Western Oregon, *Microb Ecol 39*, 12-21.

15. Poly, F., Ranjard, L., Nazaret, S., Gourbiere, F., and Monrozier, L. J. (2001) Comparison of nifH gene pools in soils and soil microenvironments with contrasting properties, *Appl Environ Microbiol 67*, 2255-2262.

16. Reiter, B., Burgmann, H., Burg, K., and Sessitsch, A. (2003) Endophytic nifH gene diversity in African sweet potato, *Can J Microbiol 49*, 549-555.

17. Zhang, Y., Dong, J., Yang, Z., Zhang, S., and Wang, Y. (2008) Phylogenetic diversity of nitrogen-fixing bacteria in mangrove sediments assessed by PCR-denaturing gradient gel electrophoresis, *Arch Microbiol 190*, 19-28.

18. Affourtit, J., Zehr, J. P., and Paerl, H. W. (2001) Distribution of Nitrogen-Fixing Microorganisms along the Neuse River Estuary, North Carolina, *Microb Ecol 41*, 114-123.

19. Zehr, J. P., Mellon, M. T., and Zani, S. (1998) New nitrogen-fixing microorganisms detected in oligotrophic oceans by amplification of Nitrogenase (nifH) genes, *Appl Environ Microbiol 64*, 3444-3450.

20. Mehta, M. P., Butterfield, D. A., and Baross, J. A. (2003) Phylogenetic diversity of nitrogenase (nifH) genes in deep-sea and hydrothermal vent environments of the Juan de Fuca Ridge, *Appl Environ Microbiol 69*, 960-970.

21. Hamilton, T. L., Boyd, E. S., and Peters, J. W. (2011) Environmental Constraints Underpin the Phylogenetic Diversity of nifH in the Yellowstone Geothermal Complex, *Microb Ecol*.

22. Mehta, M. P., and Baross, J. A. (2006) Nitrogen fixation at 92 degrees C by a hydrothermal vent archaeon, *Science 314*, 1783-1786.

23. Tyson, G. W., Lo, I., Baker, B. J., Allen, E. E., Hugenholtz, P., and Banfield, J. F. (2005) Genome-directed isolation of the key nitrogen fixer Leptospirillum ferrodiazotrophum sp. nov. from an acidophilic microbial community, *Appl Environ Microbiol 71*, 6319-6324.

24. Boyd, E. S., Jackson, R. A., Encarnacion, G., Zahn, J. A., Beard, T., Leavitt, W. D., Pi, Y., Zhang, C. L., Pearson, A., and Geesey, G. G. (2007) Isolation, characterization, and ecology of sulfur-respiring crenarchaea inhabiting acid-

sulfate-chloride-containing geothermal springs in Yellowstone National Park, *Appl Environ Microbiol 73*, 6669-6677.

25. Boyd, E. S., Cummings, D. E., and Geesey, G. G. (2007) Mineralogy influences structure and diversity of bacterial communities associated with geological substrata in a pristine aquifer, *Microb Ecol 54*, 170-182.

26. Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2007) Clustal W and Clustal X version 2.0, *Bioinformatics 23*, 2947-2948.

27. Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0, *Mol Biol Evol 24*, 1596-1599.

28. Stewart, W. D., Fitzgerald, G. P., and Burris, R. H. (1967) In situ studies on N2 fixation using the acetylene reduction technique, *Proc Natl Acad Sci U S A 58*, 2071-2078.

29. Steunou, A. S., Bhaya, D., Bateson, M. M., Melendrez, M. C., Ward, D. M., Brecht, E., Peters, J. W., Kuhl, M., and Grossman, A. R. (2006) In situ analysis of nitrogen fixation and metabolic switching in unicellular thermophilic cyanobacteria inhabiting hot spring microbial mats, *Proc Natl Acad Sci U S A 103*, 2398-2403.

30. Huelsenbeck, J. P., and Ronquist, F. (2001) MRBAYES: Bayesian inference of phylogenetic trees, *Bioinformatics 17*, 754-755.

31. Ronquist, F., and Huelsenbeck, J. P. (2003) MRBAYES 3: Bayesian phylogenetic inference under mixed models, *Bioinformatics 19*, 1572-1574.

32. Posada, D. (2006) ModelTest Server: a web-based tool for the statistical selection of models of nucleotide substitution online, *Nucleic Acids Res 34*, W700-703.

33. McCann, K. S. (2000) The diversity-stability debate, *Nature 405*, 228-233.

34. Tilman, D., Knops, J., Wedin, D., Reich, P., Ritchie, M., and Siemann, E. (1997) The influence of functional diversity and composition on ecosystem processes, *Science 277*, 1300-1302.

35. Zehr, J. P., Jenkins, B. D., Short, S. M., and Steward, G. F. (2003) Nitrogenase gene diversity and microbial community structure: a cross-system comparison, *Environ Microbiol 5*, 539-554.

36. Nordstrom, D. K., Ball, J. W., and McClesky, R. B. (2005) Ground water to surface water: chemistry of thermal outflows in Yellowstone National Park, In *Geothermal biology and geochemistry in Yellowstone National Park* (Inskeep, W. P., and McDermott, T. R., Eds.), pp 143-162, Montana State University, Bozeman.

37. Laguerre, G., van Berkum, P., Amarger, N., and Prevost, D. (1997) Genetic diversity of rhizobial symbionts isolated from legume species within the genera Astragalus, Oxytropis, and Onobrychis, *Appl Environ Microbiol 63*, 4748-4758.

38. DeWitt, T. J., Sih, A., and Wilson, D. S. (1998) Costs and limits of phenotypic plasticity, *Trends in Ecology & Evolution 13*, 77-81.

39. Dicker, H. J., and Smith, D. W. (1980) Acetylene Reduction (Nitrogen Fixation) in a Delaware, USA Salt Marsh, *Marine Biology 57*, 241-250.

40. Knowles, R. (1979) Denitrification, acetylene reduction, and methane metabolism in lake sediment exposed to acetylene, *Appl Environ Microbiol 38*, 486-493.

41. Delwiche, C. C., and Steyn, P. L. (1970) Nitrogen Isotope Fractionation in Soils and Microbial Reactions, *Environmental Science & Technology 4*, 929-935.

42. Havig, J. R., Raymond, J., Meyer-Dombard, D. R., Zolotova, N., and Shock, E. L. (2010) Merging Isotopes and Community Genomics in a Siliceous Sinter-Depositing Hot Spring, *Geochimica et Cosmochimica Atca*.

43. Wander, M. C., Kubicki, J. D., and Schoonen, M. A. (2008) Reduction of N2 by Fe2+ via homogeneous and heterogeneous reactions Part 2: the role of metal binding in activating N2 for reduction; a requirement for both pre-biotic and biological mechanisms, *Orig Life Evol Biosph 38*, 195-209.

44. Schoonen, M. A., and Xu, Y. (2001) Nitrogen reduction under hydrothermal vent conditions: implications for the prebiotic synthesis of C-H-O-N compounds, *Astrobiology 1*, 133-142.

45. Norris, P. R., Murrell, J. C., and Hinson, D. (1995) The potential for diazotrophy in iron-and sulfur-oxidizing acidophilic bacteria, *Arch. Microbiol. 164*, 294-300.

46. Estep, M. L. F., and Macko, S. A. (1984) Nitrogen isotope biogeochemistry of thermal springs, *Org. Geochem. 6*, 779-785.

CHAPTER 5

TRANSCRIPTIONAL PROFILING OF NITROGEN FIXATION IN *AZOTOBACTER VINELANDII*

Contributions of Authors and Co-Authors

Manuscript in Chapter 5

Author: Trinity L. Hamilton

Contributions: Conceived the study, collected and analyzed output data, and wrote the manuscript.

Co-author: Marcus Ludwig

Contributions: Collected and analyzed output data, provided valuable insight, discussed results and implications and edited the manuscript.

Co-author: Ray Dixon

Contributions: Discussed the results and implications, provided valuable insight, and edited the manuscript.

Co-author: Eric S. Boyd

Contributions: Discussed the results and implications, provided valuable insight, and edited the manuscript at all stages.

Co-author: Patricia Dos Santos

Contributions: Discussed the results and implications, provided valuable insight, and edited the manuscript at all stages.

Co-author: João C. Setubal

Contributions: Discussed the results and implications and edited the manuscript.

Co-author: Donald A. Bryant

Contributions: Discussed the results and implications, provided valuable insight, and edited the manuscript at all stages.

Co-author: Dennis R. Dean

Contributions: Discussed the results and implications, provided valuable insight, and edited the manuscript at all stages.

Co-author: John W. Peters

Contributions: Obtained funding, discussed the results and implications, provided valuable insight, and edited the manuscript at all stages.

Manuscript Information Page

Authors:Trinity L. Hamilton, Marcus Ludwig, Ray Dixon, Eric S. Boyd, Patricia C. Dos Santos, João C. Setubal, Donald A. Bryant, Dennis R. Dean, and John W. Peters

Journal: Journal of Bacteriology

Status of the Manuscript:
___Prepared for submission to a peer-reviewed journal
___Officially submitted to a peer-reviewed journal
___Accepted by a peer-reviewed journal
_x_Published in a peer-reviewed journal

Publisher: American Society for Microbiology

Issue in which the manuscript appears: Volume 193, 4477 – 4486 (2011)

CHAPTER 5

TRANSCRIPTIONAL PROFILING OF NITROGEN FIXATION IN *AZOTOBACTER VINELANDII*

Published: *Journal of Bacteriology*, **2011**, *193*, 4477 – 4486

Trinity L. Hamilton[1], Marcus Ludwig[2], Ray Dixon[3], Eric S. Boyd[1], Patricia C. Dos Santos[4], João C. Setubal[5,6], Donald A. Bryant[1,2], Dennis R. Dean[7], and John W. Peters[1*]

[1]*Department of Chemistry and Biochemistry and the Astrobiology Biogeocatalysis Research Center, Montana State University, Bozeman, Montana 59717*
[2]*Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, Pennsylvania 16802*
[3]*Department of Molecular Microbiology, John Innes Centre, Norwich NR4 7UH, United Kingdom*
[4]*Department of Chemistry, Wake Forest University, Winston-Salem, North Carolina 27109*
[5]*Virginia Bioinformatics Institute, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061*
[6]*Department of Computer Science, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061*
[7]*Department of Biochemistry, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061*

Abstract.

Most biological nitrogen (N_2) fixation results from the activity of a molybdenum-dependent nitrogenase, a complex iron-sulfur enzyme found associated with a diversity of bacteria and some methanogenic archaea. *Azotobacter vinelandii*, an obligate aerobe, fixes nitrogen via the oxygen-sensitive Mo-nitrogenase but is also able to fix nitrogen through the activities of genetically-distinct alternative forms of nitrogenase designated as the Vnf and Anf systems when Mo is limiting. The Vnf system appears to replace Mo by V and the Anf system is thought to contain Fe as the only transition metal within their respective active site metallocofactors. Prior genetic analyses suggest that a number of Nif-encoded components are involved in the Vnf or Anf systems. Genome-wide transcription profiling of *A. vinelandii* cultured under nitrogen-fixing conditions under varying metal amendments (e.g., Mo or V) revealed the discrete complement of genes associated with each nitrogenase system and the extent of cross-talk between the systems. In addition, changes in transcript levels of genes not directly involved in N_2 fixation

provided insight into the integration of central metabolic processes and the oxygen-sensitive process of N_2 fixation in this obligate aerobe. The results underscored significant differences between Mo-dependent and Mo-independent diazotrophic growth that highlight the significant advantages of diazotrophic growth in the presence of Mo.

Introduction

Biological nitrogen fixation, the reduction of dinitrogen (N_2) to ammonia, is an essential reaction in the global nitrogen (N) cycle. Biological N_2 fixation accounts for roughly two thirds of the fixed N produced on Earth, and is catalyzed by the nitrogenase complex (*1*). Homologs of nitrogenase have been identified in a diversity of bacteria and in several lineages of methanogenic Archaea (*2, 3*). The majority of present-day biological N_2 fixation is catalyzed by molybdenum (Mo)-nitrogenase (encoded by *nifHDK*), an oxygen-sensitive, metalloenzyme complex composed of the Fe protein (product of *nifH*) and the MoFe protein (products of *nifDK*) (*4*). The Fe protein is a homodimer bridged by an intersubunit [4Fe-4S] cluster that serves as the obligate electron donor to the MoFe protein (*5*). The MoFe protein is a $\alpha_2\beta_2$ heterotetramer that houses the P-clusters, [8Fe-7S] clusters that shuttle electrons to the FeMo-cofactors. The FeMo-cofactors are [Mo-7Fe-9S-homocitrate] clusters where substrate reduction occurs (*6*). Two genetically distinct, "alternative" forms of nitrogenase have been identified in a subset of diazotrophs that also encode for *nif* (*2, 3, 7*). The nitrogenase encoded by the *vnfHDK* genes contains vanadium in place of molybdenum in the active-site cofactor, whereas the nitrogenase encoded by the *anfHDK* genes appears to contain Fe as the only metal constituent of its active-site cofactor (*8, 9*). In some diazotrophs, the expression and activity of the alternative forms is regulated by the availability of Mo or V when fixed N

is also limiting (*10, 11*); in contrast, at least one diazotroph synthesizes the alternatives in the absence of fixed N when the Mo-dependent nitrogenase is not functional, regardless of trace metal availability (*12*). Biochemical analysis indicates that the N_2 reduction rates for the Mo-nitrogenase are much higher than the rates observed for the alternative forms (*13*). In addition, there is a larger proportion of the electron flux during nitrogen reduction are directed toward hydrogen evolution by the alternative forms of nitrogenase in comparison to Mo-nitrogenase (*14*). Although the alternative nitrogenases are genetically distinct they are mechanistically similar to the Mo-nitrogenase. Because neither the Vnf nor Anf systems contain a complete suite of genes necessary for the maturation of their corresponding catalytic partners, it is likely that significant cross-talk occurs either between or among the different systems during the assembly and regulation of these different enzyme complexes (*15, 16*).

A. vinelandii is an obligately aerobic member of the *Gammaproteobacteria* that is common in terrestrial soils sampled from across the world (*17*) and is known to encode both alternative forms of nitrogenase in addition to Nif (*10, 11, 18*). Owing to its genetic tractability, *A. vinelandii* has emerged as the principal model organism for exploring the nitrogenase mechanism, the assembly of complex metalloclusters, and the regulatory features that coordinate the differential expression of the large number of genes associated with N_2 fixation (*19*). The recently completed *A. vinelandii* genome sequence (*20*) in combination with the results of extensive biochemical and genetic characterization (*1, 21, 22*) have implicated at least eighty-two gene products in the formation and regulation of the three forms of nitrogenase (*23*). In addition to the energetic demands associated with maintaining these eighty-two genes in *A. vinelandii*,

the nitrogenase reaction is energetically demanding, requiring 16 molecules of ATP and 8 electrons for the reduction of a single molecule of N_2. These high metabolic costs, in conjunction with the demands for fixed N in growing cells and the requirement to protect the nitrogenase system from oxidative inactivation, presumably imposed a strong selective pressure to efficiently integrate N_2 fixation with other cellular metabolic processes.

In the present study, we examined patterns of global gene expression in *A. vinelandii* when cultured under N-replete conditions (fixed N source readily available) or when cells were grown under N_2-fixing conditions (fixed N source is limiting) in several experimental iterations that individually favored the expression of the *nif, vnf,* or *anf* systems. Mo-dependent diazotrophy was maintained by growing cells in medium lacking fixed N and amended with excess molybdate (Mo-dependent diazotrophy). Diazotrophic growth using the V-dependent nitrogenase was achieved by growth in medium lacking fixed N and molybdate, and amended with excess vanadate (V-dependent diazotrophy). Diazotrophic growth using the Fe-dependent nitrogenase was achieved by growth in medium lacking fixed N, molybdate, and vanadate (heterometal-independent diazotrophy). Because alternative nitrogenase dependent diazotrophic growth of *A. vinelandii* occurs only in the absence of even trace levels of Mo, RT-PCR was used to confirm the presence of transcripts of the appropriate alternative nitrogenase structural genes under the given cultivation conditions prior to transcriptome analysis. High-throughput cDNA sequencing revealed distinct patterns in transcript levels in cells cultivated under the various growth conditions examined. These results have broad

implications for the physiology of diazotrophic growth in *A. vinelandii* and the evolution of nitrogenase.

Material and Methods

Strains and Growth Conditions

Azotobacter vinelandii DJ was cultured at 30 °C with rotary shaking (200 RPM) in modified liquid Burk medium (*24*). Medium designed to favor growth under N replete conditions was amended with 25 mM filter-sterilized ammonium acetate (NH_4OAc). For alternative-nitrogenase-dependent diazotrophic growth, Mo-deficient media was prepared as previously described, and NH_4OAc was not added (*24*). For V-dependent diazotrophy, V_2O_5 was added to N- and Mo-deficient medium to a final concentration of 1 μM. Cells were grown in triplicate under each condition and nitrogenase activity was monitored by acetylene reduction (*25*). Expression of genes encoding each type of nitrogenase was confirmed by reverse transcriptase-PCR (RT-PCR) using the AccessQuick RT-PCR System, according to the manufacturer's protocol (Madison, WI). Targets and primers for RT-PCR are listed in D10 of Appendix D.

Preparation of Total RNA, cDNA Library Constructions and SOLiD™ Sequencing

After sub-culturing for at least three transfers under the specified growth conditions, cells from 30 mL cultures at an OD_{600} of 0.7 were harvested by centrifugation at 10,000 x *g* at 4 °C for 10 min. Cell pellets were flash-frozen in liquid N_2 and stored at -80 °C until further processed. Total RNA was prepared from cell pellets from each of the four conditions as previously described. Total RNA (0.5 μg) from each of the four cell-

growth conditions was submitted to the Genomics Core Facility at The Pennsylvania State University (University Park, PA) for cDNA library construction and SOLiD™ sequencing (*26*).

Data Processing and Analysis

Raw sequencing reads were mapped against the *A. vinelandii* genome as previously described (*26*). Reads that mapped to more than one region of the genome (2 to 5% of the total) could not be unambiguously mapped and were excluded from subsequent analyses (see Table D7 in Appendix D). Summary information for sequences obtained by SOLiD™ is provided in Table D7 in Appendix D. Statistical analyses were performed using *DESeq* (*27*) (version 1.2.0) from the Bioconductor R Statistical packages (http://www.r-project.org/contributors.html) (*28*). Transcript level differences with adjusted *P-values* of < 0.01 (generally, genes with more than 20 transcripts detected) and also a fold-change ratio of 2.0 or greater (see ref. (*26*)) were considered to be significant (Fig. 5.3). Previous comparisons of biological replicates have shown that two-fold changes in transcript levels are significant and that much smaller changes are probably significant for highly expressed genes (*26*). The fifteen most highly transcribed genes in each condition are summarized in Table D8 of the Appendix D.

Quantitative RT-PCR

To confirm the results of SOLiD™ sequencing, fourteen genes were chosen for qRT-PCR analyses that spanned the range of coverage depth from the four transcriptomes (See Table D9 in Appendix D). qRT-PCR was performed using the Power SYBR Green RNA-to-C_T one-step kit from Invitrogen (Carlsbad, CA) according to the manufacturer's

protocol, and reactions were assayed on a RotorGene-Q real-time PCR detection system from Qiagen (Valencia, CA). Reactions were performed in triplicate with 25 ng of total RNA quantified as above, with 200 nM forward and reverse primer (see Table D10 in Appendix D), in a final reaction volume of 20 μL. Control reactions contained either no reverse transcriptase or no template RNA. Results of the qRT-PCR assays are summarized in Fig. D1 in Appendix D.

Results and Discussion

Transcript levels for nearly 30% of the *A. vinelandii* genes changed more than two-fold during diazotrophic growth when compared to the fixed N-replete control. This included 884 genes for which the changes in expression were shared for all the diazotrophic growth conditions tested (Mo-dependent, V-dependent or heterometial-independent diazotrophy) when compared to the fixed-N replete control cultures (Fig. 5.1A). Mo-dependent diazotrophic growth resulted in the differential expression of 398 genes (at least 2-fold change compared to fixed-N replete control cultures) that did not undergo a similar change in abundance under conditions of V-dependent or heterometal independent diazotrophy (Fig. 5.1A). In addition to the 884 genes that changed under all nitrogen-fixing conditions, V- or Fe-only nitrogenase dependent diazotrophic growth resulted in an additional set of 661 common differentially expressed genes relative to the fixed-N replete control that were not observed under Mo-nitrogenase dependent diazotrophic growth. However, the number of genes uniquely affected by either V-dependent or heterometal independent diazotrophy was much smaller than for Mo-dependent growth: 161 and 175, respectively (Fig. 5.1A).

The largest changes in transcript abundance were observed during growth requiring expression of the alternative nitrogenases compared to Mo-dependent diazotrophy, when transcript abundance increased for over 1000 genes and more than 700 decreased (by at least 2-fold) when compared the fixed N-replete control (Fig. 5.1B).

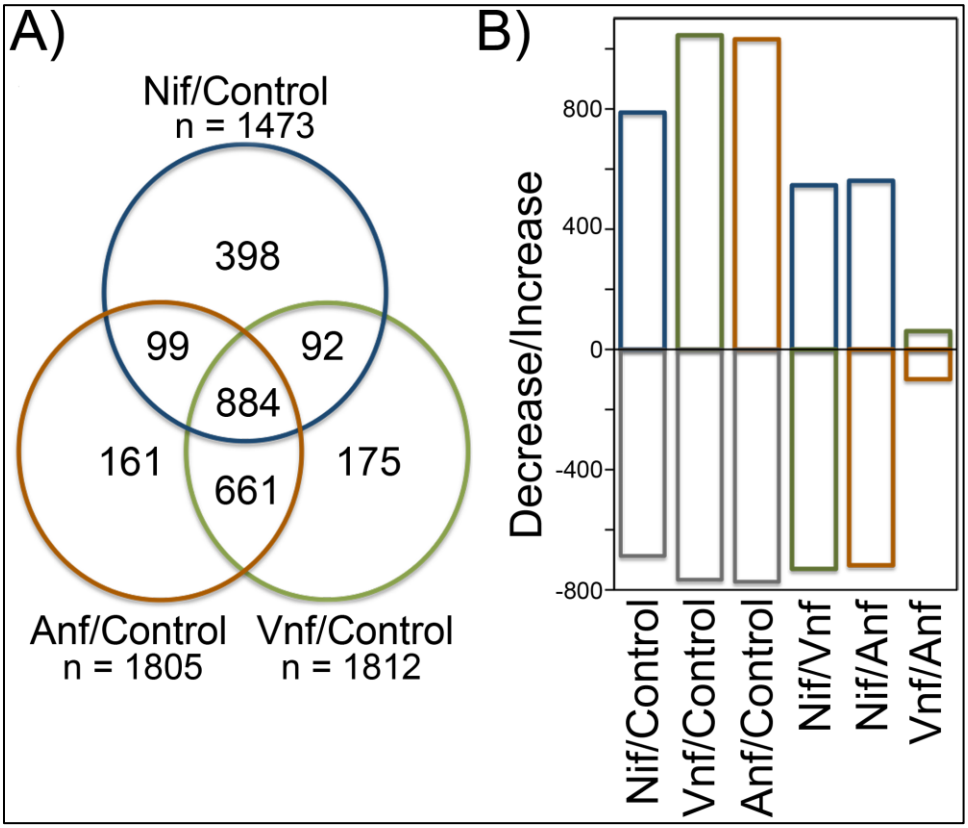


Figure 5.1. Global response of the *A. vinelandii* trancriptome during diazotrophic growth. (A) Venn diagram comparing the total number of mRNA species undergoing at least a 2.0 fold change in transcript level under Mo-dependent diazotrophy (Nif, blue circle), under V-dependent diazotrophy (Vnf, green circle), and under heterometal independent diazotrophy (Anf, rust circle) compared to a N-replete culture (Control). Numbers in overlapping circles indicate mRNA species that undergo a change in transcript level under both or all three conditions. n = the total number of mRNA species undergoing a change in transcript level in each culture. (B) Bar graph of the total number of mRNA species that undergo a larger than 2-fold increase or decrease in abundance of transcripts as a result of metal availability and nitrogen limitation.

Mo-dependent diazotrophic growth resulted in a smaller change in the transcriptome: transcripts for 788 genes increased and 685 decreased more than 2-fold (Fig. 5.1B). Interestingly, only 160 genes were differentially expressed as a result of diazotrophic growth dependent on the alternative nitrogenases when comparing V-dependent diazotrophy growth to heterometal-independent diazotrophy (Fig. 5.1B). Collectively, the results indicated that the patterns of gene expression under conditions of V- dependent or heterometal-independent diazotrophic growth are very similar to one another but distinctly different from the patterns of gene expression under Mo-dependent diazotrophic growth.

Expression of *nif* Genes Under Mo-dependent Nitrogen Fixing Conditions When Compared To Non-diazotrophic Growth

The genetic complexity of nitrogenase assembly and the regulation of *nif* gene expression of nitrogenase have been extensively characterized using the model diazotroph, *A. vinelandii* (*19*). Previous specific gene deletion (*15*, *29-31*) and northern blot-hybridization analyses (*31*, *32*) provided key insights into the genes involved in biological N_2 fixation in *A. vinelandii*. The genes encoding Mo-nitrogenase and the machinery to assemble the enzyme are co-localized in a major *nif* cluster and a secondary cluster that contains a few key biosynthetic and regulatory genes (*29, 31*) (Fig. 5.2A). Transcription of *nif* genes is activated by a σ^{54}-dependent activator, NifA, and a number of *nif*-specific promoter sequences have been identified within the two *nif* clusters (*10, 33*).

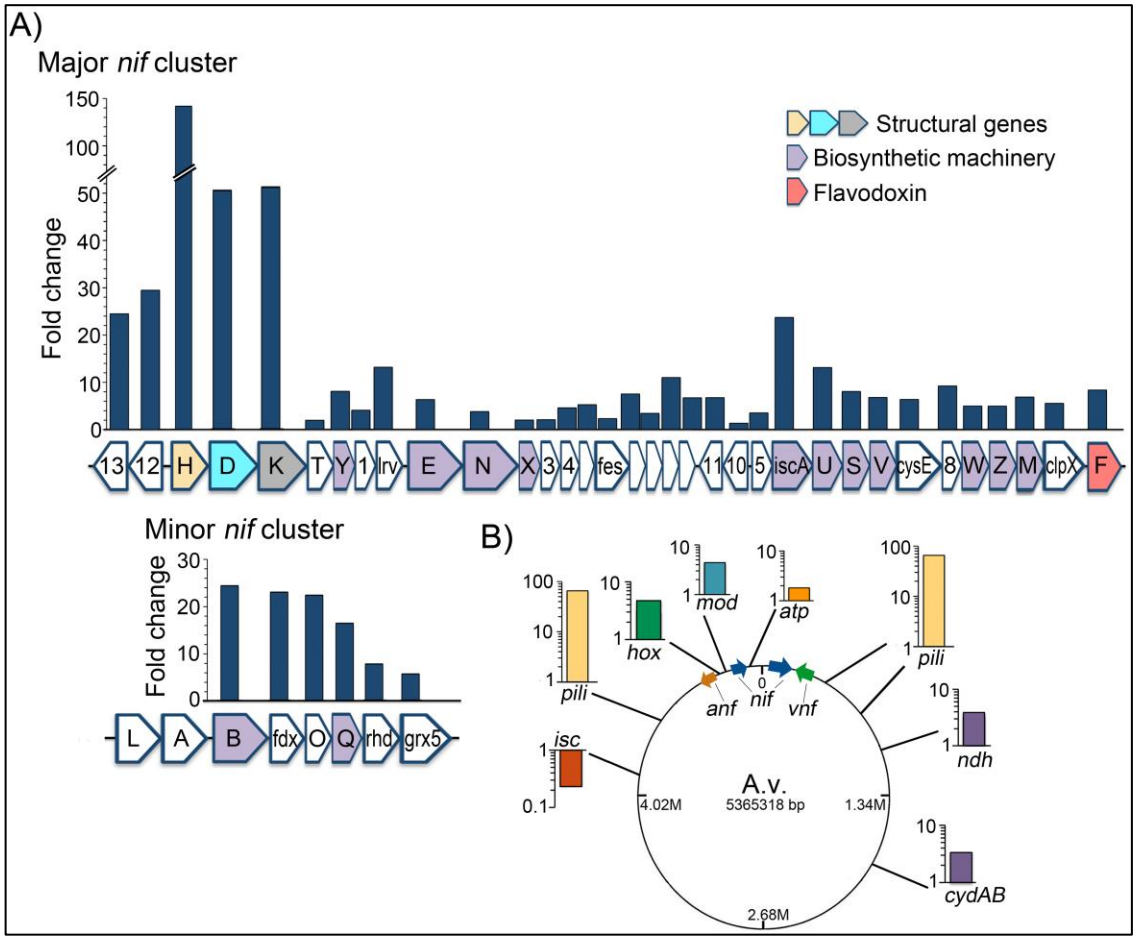


Figure 5.2. Differential expression of proteins necessary for Mo-dependent diazotrophy. (A) The ratio of transcript levels (fold change) of genes encoding structural proteins, assembly machinery and regulatory proteins for Mo-dependent diazotrophy (Nif, blue) are compared to transcript levels in fixed N replete cultures (Control). Transcripts mapping to the regulatory genes *nifA* and *nifL* were detected at very low levels in both Mo-dependent diazotrophic and fixed N replete cultures and as a result, are not included in the figure (see Table D1 in Appendix D). (B) The ratio of transcript levels of genes or groups of genes encoding proteins important for the global response of *A. vinelandii* during Mo-dependent diazotrophic growth compared to the N- replete culture (Control), presented in log scale. The genomic location of these genes are indicated as well as the genomic locations of the gene clusters encoding the Mo-dependent nitrogenase (*nif*, blue arrows) and the alternative nitrogenase (*vnf* (green arrow) and *anf* (rust arrow)). *isc*, ISC, *pili*, pilus machinery, *hox*, membrane-bound [NiFe]-hydrogenase, *mod*, molybdate transport, *atp*, ATP synthase, *ndh*, NADH dependent oxidoreductase, *cydAB*, cytochrome *bd* oxidase I. Gene names are indicated by letter(s). Full gene names, descriptions and locus tags are defined in Table D1 and Table D2 in Appendix D.

Under N-fixing conditions, nitrogenase components comprise nearly 10% of the total cellular protein, the majority of which are the structural proteins (*34*). The catalytic components necessary for FeMo-co biosynthesis are presumably necessary in much lower amounts. A large increase in transcript levels of the structural genes *nifH* (Avin01380), *nifD* (Avin01390), and *nifK* (Avin01400) was observed under Mo-dependent diazotrophy when compared to the fixed-N replete culture (control) (Fig. 5.2A, also see Table D1 in Appendix D).

Transcript levels of *nifH* (143-fold) increased much more than those of *nifD* and *nifK* (~54-fold) (Fig. 5.2A). This observation is consistent with previous analyses and supports the observation that nitrogenase activity is optimal under conditions of high Fe protein to MoFe protein ratios (*34, 35*). Interestingly, the transcript abundances for non-essential *nif* genes (*29*) downstream of *nifK* increased by only 2- to 14-fold (Fig. 5.2A). The different abundances for these apparently co-transcribed genes suggested that these gene-specific regions of the transcripts have different processing and segmental stabilities or that their synthesis may be regulated by transcriptional attenuation or processing.

The operon adjacent to the *nifHDK* structural genes contains *nifE* (Avin01450), *nifN* (Avin01470) and *nifX* (Avin01480), encoding the scaffold protein necessary for FeMo-co biosynthesis (NifEN) and a carrier protein (NifX) (*1*) (Fig. 2A, also see Table S1 in the supplemental material). The Shethna protein (FeSI, encoded by *fesI*, Avin01520), which is involved in protecting the Fe protein from O_2 damage (*36*) is also encoded within this operon. Transcripts mapping to these genes increased 2 to 6-fold under Mo-dependent nitrogen fixation conditions (Fig. 5.2A).

The majority of the remaining genes within the major *nif* cluster encode proteins important for FeMo-co and P-cluster biosynthesis and maturation of nitrogenase (*1*). Transcripts mapping to the genes *iscA*^{nif} (Avin01610) and *nifUSV* (Avin01620 to Avin01640) increased (~6- to 24-fold) under N_2-fixing conditions (Fig. 5.2A, also see Table D1 in Appendix D). NifU, a scaffold protein, and NifS, a cysteine-desulfurase, are similar to proteins of the Isc and Suf housekeeping systems for [FeS] cluster biosynthesis and facilitate the assembly of Nif-specific metal clusters (*37*). NifU and NifS are required for maturation of nitrogenase in *A. vinelandii* and despite the similarity to IscU and IscS and evidence that some cross-talk may exist (*38*), are unable to supplant the function of the ISC system (*37*). NifV is a homocitrate synthase and is required for synthesis of homocitrate for FeMo-co in *A. vinelandii* (*39*). Previous biochemical and genetic analyses indicate that these gene products are necessary for optimal diazotrophic growth in *A. vinelandii* (*29, 30*). Transcript levels for *orf8* (Avin01660) and *nifZW* (Avin01670 and Avin01680), also important for Mo-dependent diazotrophic growth (*30*); increased ~5- to 9-fold (Fig. 5.2A). Increases of ~6- to 8-fold were observed for transcripts mapping to *nifM* (Avin01690), the product of which is necessary for Fe protein maturation (*40*), and *clpX*^{nif} (Avin01700); which is co-transcribed (Fig. 2A) as well as the adjacent, independently transcribed *nifF* (Avin01710) encoding a flavodoxin (*41*) (Fig. 5.2A). Transcripts mapping to genes encoding several hypothetical proteins located within the major *nif* operon also increased 3- to 11-fold (Fig. 2A, also see Table S1 in the supplemental material). However, previous gene-deletion studies indicated that these products are not required for diazotrophic growth (*29, 30*).

The minor *nif* cluster contains two operons that include the genes encoding the *nif*-regulatory proteins *nifL* (Avin50990) and *nifA* (Avin5100) and maturation proteins encoded by *nifB* (Avin51010) and *nifQ* (Avin51040), which are involved in FeMo-co biosynthesis. The genes *nifB, fdx, nifO, nifQ, rhd* and *grx5* (Avin51010 to Avin51060) are co-transcribed. A large increase (~12- to 18-fold) was detected in the number of transcripts mapping to *nifB, fdx, nifO, nifQ*, and a smaller increase (4- to 6-fold) was noted for *rhd* and *grx5* (Fig. 5.2A). NifB is a radical SAM protein and a key enzyme required for FeMo-co biosynthesis. NifQ (Avin51040) has been implicated in Mo acquisition and FeMo-co biosynthesis (*1*). The *nifL* and *nifA* regulatory genes are located upstream of *nifB* and are under the control of a separate promoter. As noted above, the σ^{54}-dependent activator NifA is required for transcriptional activation of Mo-nitrogenase genes; NifA activity is controlled by its regulatory partner protein, NifL in response to excess oxygen and fixed N, ensuring that Mo-dependent nitrogen fixation only occurs under appropriate physiological conditions (*42*). Low transcript levels for *nifL* and *nifA* were detected under both conditions (see Table D1 in the Appendix D).

Differences in Global Patterns of Gene Expression Implicated During Mo-dependent Diazotrophic Growth When Compared To Non-diazotrophic Growth

Fixed N is essential for biosynthesis of cellular components and growth; however, when fixed N is limiting, biological N_2 fixation also imposes a high demand for ATP and reducing power on cells. These overlapping and competing requirements were expected to cause significant differences in the patterns of gene expression in the presence and absence of fixed N. Other than the major *nif* gene cluster, the largest increases (60- to >

200-fold) in transcript abundance under Mo-dependent diazotrophic growth occurred for genes that encode Type IV pili, namely *pilG*, *pilE*, *pilA*, *pilN*, and *pilM* (Avin03180, Avin11830, Avin12104, Avin45290, and Avin45300, respectively) (Fig. 5.2B, also see Table D2 in Appendix D). Type IV pili are involved in a number of processes including cell motility, DNA uptake, sensing, attachment and aggregation (55); the latter could be a previously unrecognized mechanism for O_2 protection in *A. vinelandii*.

Several strategies of O_2-protection for nitrogenase have been noted, including respiratory protection by active consumption of oxygen (*43*), spatial decoupling, for which N_2 fixation occurs in differentiated anoxic heterocysts (*44*), and temporal separation by free-living cyanobacteria that fix nitrogen at night when oxygenic photosynthesis cannot occur (*45*). When carbon is not limiting, *A. vinelandii* employs a respiratory protection mechanism to limit levels of cytoplasmic O_2, effectively protecting oxygen-sensitive enzymes such as nitrogenase from inactivation (*43*). The genome of *A. vinelandii* encodes four NADH:ubiquinone oxidoreductases and five terminal oxidases. Transcripts mapping to the genes coding for cytochrome *bd* oxidase I (CydAB I, Avin19880 and Avin19890) and the Type II NADH-dependent oxidoreductase Ndh (Avin12000) increased 3- to 4-fold in nitrogen fixing cells when compared to the control (Fig. 2B, also see Table S2 in the supplemental material). These results are consistent with previous studies indicating a role for both CydAB I and Ndh in diazotrophic growth of *A. vinelandii*, especially at high O_2 concentrations (*43*, *46*). Small increases in transcript abundance for two other terminal oxidases, cytochrome cbb_3 oxidase (Cco, Avin19940 to Avin20010) and cytochrome *c* oxidase (Cox, Avin11170 and Avin11180), were also observed during diazotrophic growth (see Table D2 in Appendix D).

The increased transcript levels for terminal oxidases and respiratory enzymes could be coordinately regulated to meet the energy demands for nitrogen fixation. Transcripts mapping to both ATP synthases (Avin19670 to Avin19750 and Avin52150 to Avin52230) (~1- to 5-fold) also increased under N_2-fixing conditions (Fig. 5.2B, also see Table S2 in the supplemental material). Transcript levels also increased ~5-fold for the well-characterized, membrane-bound [NiFe]-hydrogenase (*hoxVTRQOLMZGK*, Avin50500 to 50590) (Fig. 5.2B) (*47*), which presumably recycles the hydrogen produced as a by-product of nitrogenase-catalyzed N_2 reduction thereby capturing reducing equivalents that would otherwise be lost.

In addition to *nif*-specific genes, whose products are necessary for the initial steps in targeting Fe and S for assembly of metalloclusters associated with the maturation of nitrogenase (*nifU* and *nifS*), the genome of *A. vinelandii* also includes a suite of genes that encode proteins required for the maturation of [Fe-S] cluster-containing proteins related to other metabolic functions. The more general machinery responsible for maturation of "housekeeping" [Fe-S] proteins has been designated "ISC" (Iron-Sulfur-Cluster) components which includes *iscR, iscU, iscS, iscA, hscB hscA, fdx* and *iscX* (Avin40340 to Avin40410) (*48*). Among the ISC proteins, IscS is paralogous to NifS (38% identity), and IscU is similar in sequence (49% identity) to the N-terminal domain of NifU. Biochemical studies have established that NifS and IscS have cysteine desulfurase activity associated with the mobilization of S for [Fe-S] cluster assembly and that both IscU and the N-terminal domain of NifU can serve as *in vitro* and *in vivo* scaffolds for [Fe-S] cluster formation (*38, 49*). Genetic and physiological studies have also established that there is some functional cross-talk between the ISC and Nif systems

for [Fe-S] cluster formation (*38*). This indicates that the Nif counterparts primarily function to satisfy the increased demand for mobilization and targeting of Fe and S for [Fe-S] cluster formation associated with N_2 fixation. However the NifU and NifS and cannot entirely replace the housekeeping [Fe-S] biogenesis supported by ISC in *A. vinelandii* (*37*, *38*). The present work provides further evidence for functional cross-talk between the ISC and Nif systems because a stimulation in NifU and NifS expression under N_2 fixing conditions is accompanied by a decrease in the expression of the ISC system (Fig. 5.2B, also see Table D2 in Appendix D). This observation also extends to the expression of other housekeeping components associated with [Fe-S] protein maturation including the ErpA (Avin46030) and NfuA (Avin28760). Interestingly, ErpA has primary structure similarity to the IscA-like protein contained within the major *nif*-cluster, and NfuA has primary structure similarity when compared to the C-terminal domain of NifU. Although the specific function of these proteins has not yet been clearly established in *A. vinelandii*, current evidence indicates they could function as [Fe-S] cluster assembly scaffolds or as intermediate carriers of [Fe-S] clusters, a role that has been demonstrated for NfuA in *Synechococcus* sp. PCC 7002 (*50*). The observed ability of the Nif-specific [Fe-S] maturation components to replace the analogous housekeeping functions partially (*38*) may reflect the lower level of expression of the ISC components when cells are cultured under N_2 fixing conditions. However, expression of housekeeping [Fe-S] cluster biosynthetic proteins is subject to feedback regulation through the action of IscR, an [Fe-S] cluster containing transcription regulator. It is also possible that the respiratory protection mechanism of nitrogen fixing cells also protects

other [Fe-S] proteins from oxygen damage thus lowering the demand for [Fe-S] protein maturation.

A. vinelandii utilizes an efficient molybdate transport system, encoded by the *mod* genes (Avin50650 to Avin50690), to provide the Mo necessary for Mo-containing enzymes, including nitrogenase (*51*). The *A. vinelandii* genome contains genes encoding a high-affinity transport system (Avin50650 to Avin50690) located adjacent to the minor *nif* cluster and preceded by a putative σ^{54}-dependent promoter. Transcripts mapping to this molybdate transport system increased ~4-fold during N limitation (Fig. 5.2B, also see Table D2 in Appendix D). In addition, smaller increases in transcript abundance, (~1- to 5-fold) were also observed for the two other Mo transport systems (Avin01280 to Avin01300 and Avin50700 to 50730) (see Table D2 in Appendix D).

Transcription Profiles in Cells Expressing Alternative Nitrogenases

The alternative forms of nitrogenase were first discovered through the observation that strains of *A. vinelandii,* in which *nif* structural genes had been deleted, were still capable of diazotrophic growth in Mo-deficient, fixed N-free media (*52*). Subsequent genetic and DNA sequence analysis has revealed that genetically distinct alternative (V and Fe-only) nitrogenases are *nif* paralogs encoded within the *vnf* and *anf* clusters, which are not co-localized with either of the *nif* gene clusters (Fig. 5.2A, Fig. 5.3) (*20*). The expression of the alternative systems in *A. vinelandii* is regulated by the availability of Mo (*32*). Transcripts of the *vnf* and *anf* structural genes have been observed by Northern blot-hybridization in N-limited *A. vinelandii* cultures in the absence of Mo with added V (*53*) or in the absence of both Mo and V (*54*).

The alternative forms of nitrogenase are only present in the genomes of organisms that also harbor the *nif* system (*2*, *3*, *7*), and although the structural genes are distinct, the gene operons encoding the alternative enzymes do not include paralogs for all of the genes for the biosynthetic machinery necessary for assembly and maturation of the Mo-dependent nitrogenase. In *A. vinelandii*, limited biochemical evidence has suggested that VnfH can serve an analogous role to NifH in FeMo-co maturation (*55*); however diazotrophic growth is not supported in mutants containing deletions of each of their respective structural genes. Moreover, mutational analyses have indicated that some components of the *nif*-encoded biosynthetic machinery, such as *nifU*, *nifS*, *nifV*, *nifM*, and *nifB*, are required for nitrogen fixation by cells utilizing either of the alternative enzymes in *A. vinelandii* (*15*, *16*) and a role for the NifEN scaffold protein has also been observed (*56*).

The transcriptome of cultures the relying on the V-dependent nitrogenase for diazotrophic growth revealed a large increase (68- to 137-fold) in the *vnfHF and vnfDGK* genes, compared to fixed N-replete cultures (control) (Fig. 5.3A). The large difference in transcript abundance observed for *nifH* versus *nifDK*, which are co-transcribed from a single promoter, was not observed for *vnfH*, *vnfD*, and *vnfK*, which are located in separate operons (*57*). A small increase (~2 fold) was observed for *nifH* transcripts but given the high transcript abundance of *vnfH*, it is unclear what the potential role of *nifH* encoded Fe protein might be during V-dependent diazotrophy. These genes encode an Fe protein specific to the V-dependent system (encoded by *vnfH*) and the VFe protein (encoded by *vnfDGK*) analogous to the MoFe protein which houses the active site metalloclusters, the FeV-cos. The *vnf* gene cluster in *A. vinelandii* also encodes a putative FeV-co assembly

scaffold, VnfEN (Avin02770 and Avin02750), paralogous to NifEN scaffold protein necessary for FeMo-co biosynthesis. Mutant strains of *A. vinelandii* with knockouts in *vnfEN* can still grow diazotrophically under Mo-limiting but V-replete diazotrophic growth conditions (*56*). However, the transcriptome analysis revealed a large increase (26- to 56-fold) in the expression of *vnfE* and *vnfN* transcripts (Fig. 5.3) and a very small increase in the expression of *nifEN* transcripts (~1- to 2.5-fold, see Table D3 in the Appendix D) during V-dependent diazotrophy.

The large increase in transcript levels for *vnfEN* supports the limited biochemical evidence that suggested that VnfEN, and not NifEN, is the preferred scaffold for both VFe-co maturation (*56*). The *vnf* operon also encodes VnfX (Avin02740) and VnfY (Avin02570) (Fig. 5.2A), paralogs of NifX and NifY. VnfY is required for optimal V-dependent diazotrophy (*58*) and large increases in transcript levels of *vnfX* and *vnfY* (15- and 117-fold) were observed during V-dependent diazotrophy (Fig. 5.3A). Interestingly, this is markedly different to that observed for the magnitude of transcript levels of *nifX* (Avin01480) and *nifY* under Mo-dependent diazotrophy, for which much smaller increases (<8-fold) were observed (see Table D1 and D3 in Appendix D). This implies that perhaps VnfX and VnfY have a more significant role in FeV-co biosynthesis than NifX and NifY in FeMo-co biosynthesis.

Transcripts mapping to the *nif*-encoded cluster biosynthesis machinery implicated in playing a role in assembly of the alternative cofactors such as *nifUSV*, *nifM*, and *nifB*, also increased (5- to 10-fold) consistent with what was observed from previous deletion mutant analysis (*15*) indicating these biosynthetic genes are also involved in V-dependent diazotrophy.

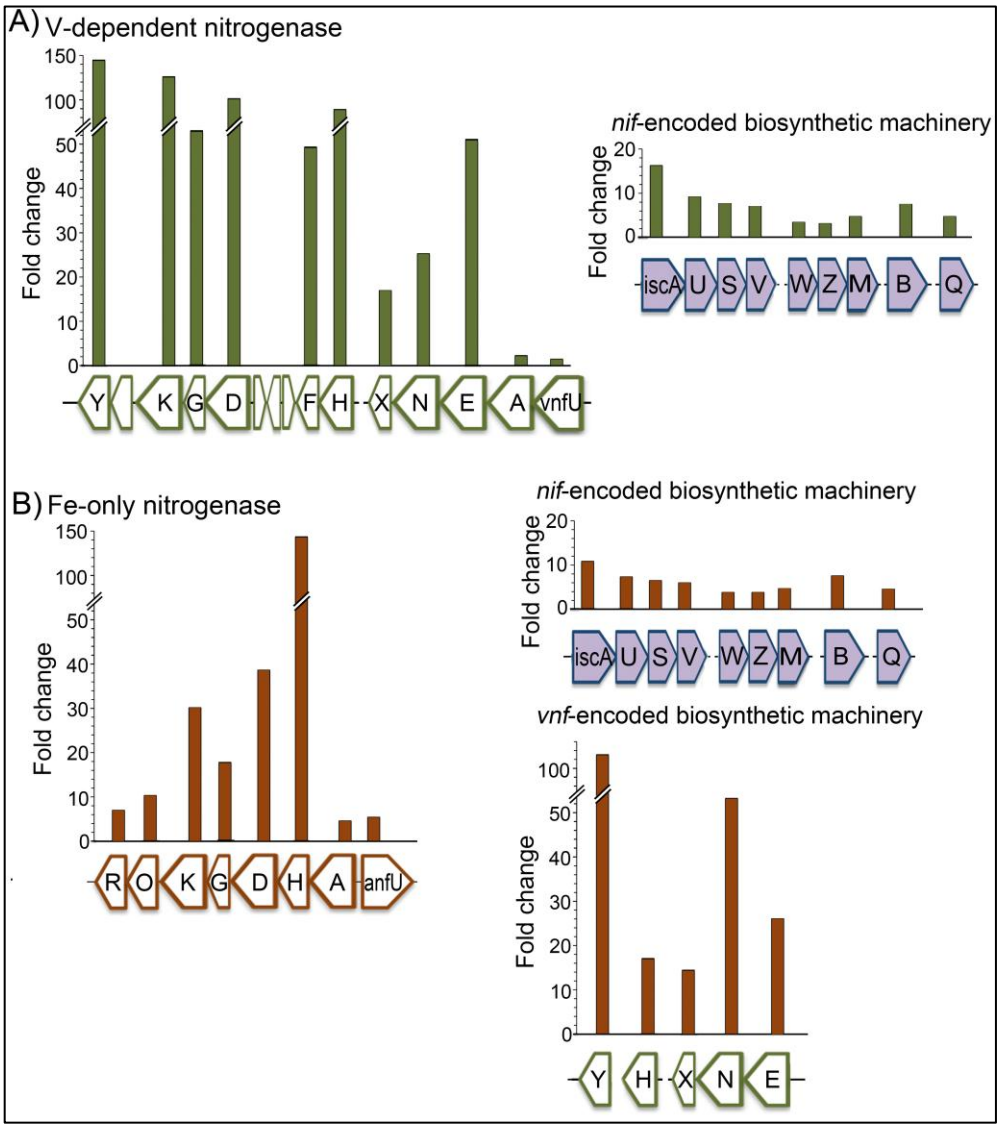


Figure 5.3. Differential expression of proteins necessary for Mo-independent nitrogen fixation. (A) The ratio of transcript levels (fold change) of genes encoding structural proteins, assembly machinery and regulatory proteins during V-dependent diazotrophy (Vnf, green) are compared to transcript levels of the fixed N replete cultures (Control). Transcript levels of *nif*-encoded genes observed to undergo large fold changes for V-dependent diazotrophy (Vnf, green) compared to transcript levels of the fixed N replete cultures (Control) are also indicated. (B) The ratio of transcript levels (fold change) of genes encoding structural proteins, assembly machinery and regulatory proteins for heterometal independent diazotrophy (Anf, rust) are compared to transcript levels of the fixed N replete cultures (Control). Transcript levels of *nif*- and *vnf*-encoded genes observed to undergo large fold changes during heterometal-independent diazotrophy (Anf, rust) compared to transcript levels of the fixed N replete cultures (Control) are also indicated. Gene names are indicated by letter(s). Full gene names, descriptions and locus tags are defined in Table S3 of the supplemental material.

Interestingly, an increase in transcript abundance of *nifQ*, necessary for Mo acquisition for FeMo-co, was also observed, and this may reflect the pressure to acquire Mo to utilize the Mo-dependent nitrogenase preferentially or simply the result of the transcriptional coupling of *nifQ* with *nifB*. Our analysis thus implicated the involvement of *nif*-encoded maturation proteins in V-dependent diazotrophy in *A. vinelandii*.

The Fe-only nitrogenase (encoded AnfHDGK; Avin49000 to Avin48970), responsible for heterometal-independent diazotrophy, is encoded by the *anf* gene cluster. This cluster includes *anfHDGK* structural genes encoding an additional Fe protein, AnfH, specific to the Fe-only system, and the genes encoding the Fe-only version of the substrate reduction component, the FeFe protein (AnfDGK), where the FeFe-cos reside. In addition, this cluster includes *anfA* (Avin49020), a regulatory gene, and *anfO* (Avin48960) and *anfR* (Avin48950) (Fig. 5.3B). Transcripts mapping to the *anfHDGK* structural genes increased 18- to 144-fold in cultures dependent on the Fe-only nitrogenase for diazotrophic growth compared to the fixed N-replete cultures (Fig. 5.3B). The differences in transcript abundance of *anfH* versus *anfD* and *anfK*, which are presumably co-transcribed, mirrors what is observed for the *nif* structural genes where increased transcript abundance for *nifH* was much larger than the increases observed for *nifD* and *nifK* transcripts. Large increases (14- to 117-fold) in the *vnf*-encoded *H*, *E*, *N*, *X*, and *Y* transcripts were observed during heterometal-independent diazotrophy (Fig. 5.3B). This observation suggested that these *vnf*-encoded products are preferred over their *nif*-encoded counterparts for maturation of the Fe-only nitrogenase. Therefore, both the V-dependent and Fe-only nitrogenases apparently utilize the VnfEN scaffold protein for cofactor assembly but the *nif*-encoded accessory proteins, NifU, NifS, NifM, NifV and

NifB for maturation (*16*). Importantly, *vnfEN* in *A. vinelandii* is derived from *nifEN*, suggesting that the preferential utilization of the VnfEN scaffold over NifEN scaffold in alternative systems is a recent evolutionary feature (*2*). In addition, VnfH has been demonstrated to be necessary for transcription of *anfHDGK*; however, in VnfH mutants, the presence of NifH is detected but lower growth rates of heterometal-independent diazotrophically grown cells are observed (*59*). An increase in *vnfH* transcripts (~17-fold) and *nifH* transcripts (~2.5-fold) were observed during heterometal-independent diazotrophy (see Table D3 in Appendix D). Together, these results may suggest that Vnf preceded and is evolutionarily ancestral to Anf, which would be consistent with evidence reported previously (*2*, *3*).

Electron Transport for Mo-, V-, and Fe-only Nitrogenase Dependent Diazotrophic Growth

The energy and electron transport requirements differ for the three forms of nitrogenase, and both the *nif* and *vnf* systems encode specific flavodoxins, *nifF* (Avin01710) and *vnfF* (Avin02650). Although *nifF* is not required for Mo-dependent diazotrophy (*41*), it is not known if *vnfF* is required for V-dependent or heterometal-independent diazotrophy. Both *vnfF* and *nifF* transcripts increased under Mo-independent diazotrophic conditions, and the increases for *vnfF* (47- to 102-fold) were much larger than for *nifF* (~4-fold) (Fig. 5.3). During Mo-dependent diazotrophy, transcripts mapping to *nifF* increased 8-fold.

Three clusters of genes, *rnf1* (Rnf1, Avin50930 to 50980), *rnf2* (Rnf2, Avin19220 to Avin19270), and *fix* (Avin10510 to Avin10550), encode electron transport systems that are thought to provide reducing equivalents to nitrogenase in *A. vinelandii* (*60*, *61*).

The *rnf* genes were originally identified and characterized in *Rhodobacter capsulatus*, and they encode a membrane-bound complex involved in electron transport to nitrogenase (*60*). The expression of the genes encoding the Rnf1 complex is NifA–dependent while the genes encoding Rnf2 are not regulated as a result of N availability (see Table D4 in Appendix D) (*62*). In *A. vinelandii*, the *rnf* genes are necessary for full Mo-dependent nitrogenase activity and have been implicated to be regulated by NifA, potentially by modulating the redox state of NifL (*62*). In addition, Rnf1 proteins are required for accumulation of the [4Fe-4S] cluster in the *nifH* encoded Fe protein (*62*). The *fix* operon, which is preceded by a putative σ^{54}-dependent promoter (R. D., unpublished data), encodes five proteins, FixP, A, B, C, and X, that form an electron transfer complex that seems to be required to support nitrogen fixation in some diazotrophs (*61*). Transcript levels for both *rnf1* and *fixPABCX* increased markedly during diazotrophic growth; however, during V-dependent or heterometal-independent diazotrophy, transcript levels for the *fixPABCX* operon were higher than those for the *rnf1* operon (~231-fold vs. ~18-fold) (see Table D4 in Appendix D). In contrast, the increase in transcripts from the *rnf1* operon and *fixPABCX* operon were of similar magnitude during Mo-dependent diazotrophy (~40-fold). As would be expected, transcript levels from the *Rnf2* operon, which is not NifA-regulated, were relatively unchanged.

Regulation of Gene Expression During Diazotrophic Growth

The σ^{54}-dependent activator proteins, VnfA and AnfA, are required for expression of V-nitrogenase and the Fe-only nitrogenase, respectively (*63*), but unlike NifA, their

activity is not subject to regulation by a NifL-like protein. As observed for *nifA*, the *vnfA* and *anfA* transcripts were detectable at low levels under all growth conditions tested here; however, their transcript levels increased under V-dependent and heterometal-independent diazotrophic growth (Fig. 5.3) and decreased in the presence of Mo (see Table D5 in Appendix D). This pattern of regulation of *vnfA* and *anfA* expression has been observed previously, and the Mo-responsive transcriptional regulator, ModE, has been shown to be at least partially involved in Mo-dependent repression of *anfA* (*63*). Transcriptional regulation of *vnfA* and *anfA* is likely to be responsible for the absence of the VFe- and Fe-only nitrogenases when molybdenum is available, because transcription of the genes encoding these alternative enzymes cannot be activated in the absence of VnfA or AnfA, respectively (*64*). The *A. vinelandii* genome contains an additional gene homologous to *nifA*, *nifA2* (Avin26490), and two additional *vnfA* homologs, *vnfA2* (Avin33440) and *vnfA3* (Avin47100) (*20*). Interestingly, the pattern of regulation of *vnfA2* and *vnfA3* expression was similar to that of *vnfA* and *anfA*, as transcripts for both increased under Mo-deplete diazotrophic conditions compared with cultures grown with molybdate (see Table D5 in Appendix D). Although the precise role of these activator paralogs is unclear at present, the sequence similarity of their DNA binding determinants to those of the cognate NifA and VnfA activators (*20*), suggests that they may be able to provide additional regulatory input.

Differences in Global Patterns of Gene Expression Implicated During Mo-independent Nitrogen Fixing Growth

The complexity of the nitrogenase enzyme, and the absolute requirement for fixed N to meet cellular growth, might be expected to result in a similar global response to

fixed-N limitation, regardless of metal availability, but as noted above, the patterns of gene expression for Mo-dependent diazotrophy are significantly different when compared to cells grown under Mo-independent diazotrophic conditions. The alternative forms of nitrogenase are markedly less efficient than the Mo-nitrogenase (*13*), and coupled with the demand for fixed N levels necessary to maintain cell growth, this inefficiency would presumably result in significant differences in the transcriptional response of V-dependent or heterometal-independent diazotrophy compared to N-replete control cultures. Key differences are observed in the *A. vinelandii* transctriptome in response to N-limiting conditions and the availability of Mo. Most notably, transcripts for *hutU* (Avin26180), the gene encoding urocanate hydratase, increased dramatically (30- to 40-fold) under diazotrophic conditions in the absence of Mo (see Table D5 in Appendix D). Urocanate hydratase catalyzes a key step in the degradation of histidine (*65*). The increased transcript levels for *hutU* presumably reflect the low catalytic efficiency of the alternative forms and imply that cells seek to supplement the fixed-N provided by the alternative forms by liberating fixed N during protein turnover by His degradation to maintain cell viability until Mo or fixed N becomes available. This increase in enzymes capable of accessing fixed N was also noted in the transcriptome analysis of *Rhodospeudomonas palustris* during diazotrophic conditions requiring expression of the alternative nitrogenases (*12*). Transcript levels of other *hut* genes within in the same operon, *hutG* (Avin26150), *hutI* (Avin26160), and a hypothetical protein (Avin26170), and just downstream, *hutH* (Avin26290) and *hutF* (Avin26300), were also observed to increase (~5- to 40-fold) under these conditions (see Table D5 in Appendix D). Interestingly, the large increase in transcripts for the formation of pili was not observed

during Mo-independent, diazotrophic growth. Also notable was an increase in the levels of transcripts for the genes (Avin04360 to Avin04410) encoding an uncharacterized soluble [NiFe]-hydrogenase, which increased ~10-fold during V-dependent and ~20-fold during heterometal-independent diazotrophy (see Table D5 in Appendix D). The increase in transcripts may reflect the need to recycle the larger amount of hydrogen produced by the activity of the alternative enzymes to counteract the catalytic inefficiency. The alternative nitrogenases have been shown to produce more hydrogen during dinitrogen reduction which presumably provides additional pressure to avoid the loss of needed reducing equivalents (*13*). Finally, transcripts mapping to a number of genes annotated as hypothetical proteins were also observed to undergo large changes as a result of Mo-independent diazotrophy. For example, Avin02580, Avin16830, Avin47120, Avin46150, and Avin10510 transcript levels increased and Avin22850, Avin09340, Avin49410, and Avin47230 transcript levels decreased under these conditions (see Table D6 in Appendix D).

Although our analysis indicates similar patterns of gene expression when cultures are grown either under conditions of V-dependent diazotrophy and heterometal-independent diazotrophy, several key differences were noted. Importantly, transcripts mapping to several genes near the *vnf* operon increased dramatically (53- to 100-fold change) during V-dependent diazotrophy (see Table D6 in Appendix D). These genes, Avin02540, Avin02550, and Avin02560, encode proteins that are homologous to the inner membrane transport protein and substrate binding protein of an ABC transporter, which is annotated as a phosphonate transport system. Due to the proximity of these genes to the *vnf* gene cluster, and to the large increase in transcripts mapping to these

genes during V-dependent diazotrophy, we hypothesize that these gene products are responsible for vanadate uptake.

Like Mo-dependent diazotrophy, both V-dependent and heterometal independent diazotrophic growth resulted in increased transcript abundance of CydAB I (~3- to 5-fold), Ndh (~3-fold) and both ATP synthases (see Table D5 in Appendix D) compared to N-replete control cultures. Likewise, similar increases in levels transcripts mapping to the *hox*-encoded membrane bound [NiFe]-hydrogenase and the Mod transport systems were also observed in cultures expressing the alternative nitrogenases. Interestingly, a decrease in the *isc* machinery was also noted, further supporting the hypothesis that *isc* gene expression levels observed under normal growth conditions are sufficient to provide [Fe-S] clusters for general cell metabolism (see Table D5 in Appendix D). A significant decrease in transcript abundance was noted for some genes regardless of metal availability during diazotrophic growth (see Table D6 in Appendix D). The differential expression of these proteins presumably reflects changes in the transcriptome of *A. vinelandii* in response to fixed-N limitation; however, most of these transcripts mapped to proteins without defined functions (for example, Avin9340, Avin22850, Avin47230, and Avin49410).

Evolutionary Implications

These results presented herein have fundamental implications on our understanding *nif* evolution. Several phylogenetic studies have suggested that the alternative nitrogenases are evolutionary ancestors of Mo-nitrogenase (*2, 3, 7*). This is considered to be consistent with early Earth history, because the bioavailability of Mo

and perhaps V would have been very low prior to the advent of oxygenic photosynthesis and subsequent oxygenation of Earth's atmosphere (*66*). Interestingly, however, the V- and Fe-only nitrogenases in extant organisms have not yet been observed in a background devoid of the Mo-nitrogenase (*2, 3*). In this study, we clearly show that FeMo-co biosynthetic genes clustered with and required for the *nif* system are expressed as presumably essential components of alternative systems, extending mutant and biochemical analyses that have suggested this functional cross-talk (*15, 16*) and in line with previous genetic experiments which indicated that the alternative nitrogenases are not functional in the absence of these *nif*-encoded gene products. The available information regarding alternative nitrogenase genomic occurrence, the results of previous genetic studies and the transcription profiling results described here collectively make it difficult to consider the alternative nitrogenases as ancestral forms of the Mo-nitrogenase. Instead a simpler hypothesis would be that the alternative nitrogenases evolved from Mo-nitrogenase to expand the fundamental niche of diazotrophic populations under conditions where Mo may in fact be limiting.

Conclusion

The carefully controlled transcriptional analysis presented here has clearly defined the suite of genes associated with each of the three nitrogenase systems in *A. vinelandii*. In addition, these results have uncovered several interesting surprises regarding the evolutionary relationships between Mo-nitrogenase and the alternative nitrogenases, which can be examined further through combined genetic, biochemical, and evolutionary analyses. The results of the global expression under N_2 fixing conditions

suggest that Mo-limitation, whether in the presence or absence of V, results in marked differences in gene expression in *A. vinelandii*. These results also support the hypothesis that the lower catalytic efficiencies of the V-dependent and Fe-only nitrogenases impose significant physiological constraints upon *A.vinelandii* when grown under N_2 fixing conditions in the absence of Mo. These studies will help to refine our understanding of the complex metabolic and regulatory interactions that control N_2 fixation in this important model organism.

Acknowledgments

This work was supported by the NASA Astrobiology Institute grant NNA08C-N85A to J.W.P. D.A.B. acknowledges support form NASA Astrobiology: Exobiology and Evolutionary Biology, Award NNX09AM87G and work in the laboratory of D.R.D is supported by NSF MCB-071770. T.L.H. was supported by an NSF-Integrated Graduate Educational Research and Training fellowship grant and E.S.B. was supported by a fellowship from the NASA Astrobiology Institute Postdoctoral Program.

Supporting Information Available: Supplemental tables of the changes in transcript levels of *nif*-specific genes, genes implicated in the global response for Mo-dependent nitrogen fixation, *nif*-, *vnf*-, and *anf*-specific genes, σ^{54}-dependent activators, *rnf*, and *rnf1* operons and the *fix* operon, the changes in transcript abundance of the top 10 up- and down-regulated genes under each growth conditions, the 15 most abundant mRNA species under each condition, transcript counts for differentially expressed genes chosen for RT-PCR and qRT-PCR analyses and one figure of the results of qRT-PCR analysis.

This material is available free of charge via the Internet at http://jb.asm.org. The relevant supplemental information is located in Appendix D.

Literature Cited

1. Rubio, L. M., and Ludden, P. W. (2008) Biosynthesis of the iron-molybdenum cofactor of nitrogenase, *Annu Rev Microbiol 62*, 93-111.

2. Boyd, E. S., Anbar, A. D., Miller, S., Hamilton, T. L., Lavin, M., and Peters, J. W. (2011) A Late Methanogen Origin for Molybdenum-Dependent Nitrogenase, *Geiobiol*.

3. Raymond, J., Siefert, J. L., Staples, C. R., and Blankenship, R. E. (2004) The natural history of nitrogen fixation, *Mol Biol Evol 21*, 541-554.

4. Burgess, B. K., and Lowe, D. J. (1996) Mechanism of Molybdenum Nitrogenase, *Chem Rev 96*, 2983-3012.

5. Georgiadis, M. M., Komiya, H., Chakrabarti, P., Woo, D., Kornuc, J. J., and Rees, D. C. (1992) Crystallographic structure of the nitrogenase iron protein from Azotobacter vinelandii, *Science 257*, 1653-1659.

6. Kim, J., and Rees, D. C. (1992) Structural models for the metal centers in the nitrogenase molybdenum-iron protein, *Science 257*, 1677-1682.

7. Soboh, B., Boyd, E. S., Zhao, D., Peters, J. W., and Rubio, L. M. (2011) Substrate specificity and evolutionary implications of a NifDK enzyme carrying NifB-co at its active site, *FEBS Lett 584*, 1487-1492.

8. Chisnell, J. R., Premakumar, R., and Bishop, P. E. (1988) Purification of a second alternative nitrogenase from a nifHDK deletion strain of Azotobacter vinelandii, *J Bacteriol 170*, 27-33.

9. Hales, B. J., Case, E. E., Morningstar, J. E., Dzeda, M. F., and Mauterer, L. A. (1986) Isolation of a new vanadium-containing nitrogenase from Azotobacter vinelandii, *Biochemistry 25*, 7251-7255.

10. Joerger, R. D., Jacobson, M. R., and Bishop, P. E. (1989) Two nifA-like genes required for expression of alternative nitrogenases by Azotobacter vinelandii, *J Bacteriol 171*, 3258-3267.

11. Joerger, R. D., Loveless, T. M., Pau, R. N., Mitchenall, L. A., Simon, B. H., and Bishop, P. E. (1990) Nucleotide sequences and mutational analysis of the structural genes for nitrogenase 2 of Azotobacter vinelandii, *J Bacteriol 172*, 3400-3408.

12. Oda, Y., Samanta, S. K., Rey, F. E., Wu, L., Liu, X., Yan, T., Zhou, J., and Harwood, C. S. (2005) Functional genomic analysis of three nitrogenase isozymes

in the photosynthetic bacterium Rhodopseudomonas palustris, *J Bacteriol 187*, 7784-7794.

13. Eady, R. R. (1996) Structureminus signFunction Relationships of Alternative Nitrogenases, *Chem Rev 96*, 3013-3030.

14. Scott, D. J., Dean, D. R., and Newton, W. E. (1992) Nitrogenase-catalyzed ethane production and CO-sensitive hydrogen evolution from MoFe proteins having amino acid substitutions in an alpha-subunit FeMo cofactor-binding domain, *J Biol Chem 267*, 20002-20010.

15. Joerger, R. D., Premakumar, R., and Bishop, P. E. (1986) Tn5-induced mutants of Azotobacter vinelandii affected in nitrogen fixation under Mo-deficient and Mo-sufficient conditions, *J Bacteriol 168*, 673-682.

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

17. Kennedy, C., Rudnick, P., MacDonald, T., and Melton, T. (2005) Genus Azotobacter, In *Bergey's manual of sytematic bacteriology* (Garrity, G. M., Ed.), pp 384-401, Spring-Verlag, New York, NY.

18. Joerger, R. D., and Bishop, P. E. (1988) Bacterial alternative nitrogen fixation systems, *Crit Rev Microbiol 16*, 1-14.

19. Kennedy, C., and Toukdarian, A. (1987) Genetics of azotobacters: applications to nitrogen fixation and related aspects of metabolism, *Annu Rev Microbiol 41*, 227-258.

20. Setubal, J. C., dos Santos, P., Goldman, B. S., Ertesvag, H., Espin, G., Rubio, L. M., Valla, S., Almeida, N. F., Balasubramanian, D., Cromes, L., Curatti, L., Du, Z., Godsy, E., Goodner, B., Hellner-Burris, K., Hernandez, J. A., Houmiel, K., Imperial, J., Kennedy, C., Larson, T. J., Latreille, P., Ligon, L. S., Lu, J., Maerk, M., Miller, N. M., Norton, S., O'Carroll, I. P., Paulsen, I., Raulfs, E. C., Roemer, R., Rosser, J., Segura, D., Slater, S., Stricklin, S. L., Studholme, D. J., Sun, J., Viana, C. J., Wallin, E., Wang, B., Wheeler, C., Zhu, H., Dean, D. R., Dixon, R., and Wood, D. (2009) Genome sequence of Azotobacter vinelandii, an obligate aerobe specialized to support diverse anaerobic metabolic processes, *J Bacteriol 191*, 4534-4545.

21. Seefeldt, L. C., Hoffman, B. M., and Dean, D. R. (2009) Mechanism of Mo-dependent nitrogenase, *Annu Rev Biochem 78*, 701-722.

22. Peters, J. W., Fisher, K., and Dean, D. R. (1995) Nitrogenase structure and function: a biochemical-genetic perspective, *Annu Rev Microbiol 49*, 335-366.

23. Dos Santos, P. C., and Dean, D. R. (2011) Coordination and fine-tuning of nitrogen fixation in Azotobacter vinelandii, *Molecular Microbiology*.

24. Bishop, P. E., Jarlenski, D. M., and Hetherington, D. R. (1982) Expression of an alternative nitrogen fixation system in Azotobacter vinelandii, *J Bacteriol 150*, 1244-1251.

25. Hardy, R. W., Holsten, R. D., Jackson, E. K., and Burns, R. C. (1968) The acetylene-ethylene assay for n(2) fixation: laboratory and field evaluation, *Plant Physiol 43*, 1185-1207.

26. Ludwig, M., and Bryant, D. A. (2011) Transcription Profiling of the Model Cyanobacterium Synechococcus sp. Strain PCC 7002 by Next-Gen (SOLiD) Sequencing of cDNA, *Frontiers in microbiology 2*, 41.

27. Anders, S., and Huber, W. Differential expression analysis for sequence count data, *Genome Biol 11*, R106.

28. Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., Hornik, K., Hothorn, T., Huber, W., Iacus, S., Irizarry, R., Leisch, F., Li, C., Maechler, M., Rossini, A. J., Sawitzki, G., Smith, C., Smyth, G., Tierney, L., Yang, J. Y., and Zhang, J. (2004) Bioconductor: open software development for computational biology and bioinformatics, *Genome Biol 5*, R80.

29. Jacobson, M. R., Brigle, K. E., Bennett, L. T., Setterquist, R. A., Wilson, M. S., Cash, V. L., Beynon, J., Newton, W. E., and Dean, D. R. (1989) Physical and genetic map of the major nif gene cluster from Azotobacter vinelandii, *J Bacteriol 171*, 1017-1027.

30. Jacobson, M. R., Cash, V. L., Weiss, M. C., Laird, N. F., Newton, W. E., and Dean, D. R. (1989) Biochemical and genetic analysis of the nifUSVWZM cluster from Azotobacter vinelandii, *Mol Gen Genet 219*, 49-57.

31. Joerger, R. D., and Bishop, P. E. (1988) Nucleotide sequence and genetic analysis of the nifB-nifQ region from Azotobacter vinelandii, *J Bacteriol 170*, 1475-1487.

32. Jacobson, M. R., Premakumar, R., and Bishop, P. E. (1986) Transcriptional regulation of nitrogen fixation by molybdenum in Azotobacter vinelandii, *J Bacteriol 167*, 480-486.

33. Dixon, R. (1998) The oxygen-responsive NIFL-NIFA complex: a novel two-component regulatory system controlling nitrogenase synthesis in gamma-proteobacteria, *Arch Microbiol 169*, 371-380.

34. Dingler, C., Kuhla, J., Wassink, H., and Oelze, J. (1988) Levels and activities of nitrogenase proteins in Azotobacter vinelandii grown at different dissolved oxygen concentrations, *J Bacteriol 170*, 2148-2152.

35. Dilworth, M. J., Subramanian, D., Munson, T. O., and Burris, R. H. (1965) The adenosine triphosphate requirement for nitrogen fixation in cell-free extracts of Clostridium pasteurianum, *Biochim Biophys Acta 99*, 486-503.

36. Maier, R. J., and Moshiri, F. (2000) Role of the Azotobacter vinelandii nitrogenase-protective shethna protein in preventing oxygen-mediated cell death, *J Bacteriol 182*, 3854-3857.

37. Johnson, D. C., Dos Santos, P. C., and Dean, D. R. (2005) NifU and NifS are required for the maturation of nitrogenase and cannot replace the function of isc-gene products in Azotobacter vinelandii, *Biochem Soc Trans 33*, 90-93.

38. Dos Santos, P. C., Johnson, D. C., Ragle, B. E., Unciuleac, M. C., and Dean, D. R. (2007) Controlled expression of nif and isc iron-sulfur protein maturation components reveals target specificity and limited functional replacement between the two systems, *J Bacteriol 189*, 2854-2862.

39. Zheng, L., White, R. H., and Dean, D. R. (1997) Purification of the Azotobacter vinelandii nifV-encoded homocitrate synthase, *J Bacteriol 179*, 5963-5966.

40. Howard, K. S., McLean, P. A., Hansen, F. B., Lemley, P. V., Koblan, K. S., and Orme-Johnson, W. H. (1986) Klebsiella pneumoniae nifM gene product is required for stabilization and activation of nitrogenase iron protein in Escherichia coli, *J Biol Chem 261*, 772-778.

41. Bennett, L. T., Jacobson, M. R., and Dean, D. R. (1988) Isolation, sequencing, and mutagenesis of the nifF gene encoding flavodoxin from Azotobacter vinelandii, *J Biol Chem 263*, 1364-1369.

42. Dixon, R., and Kahn, D. (2004) Genetic regulation of biological nitrogen fixation, *Nat Rev Microbiol 2*, 621-631.

43. Poole, R. K., and Hill, S. (1997) Respiratory protection of nitrogenase activity in Azotobacter vinelandii--roles of the terminal oxidases, *Biosci Rep 17*, 303-317.

44. Fay, P. (1992) Oxygen relations of nitrogen fixation in cyanobacteria, *Microbiol Rev 56*, 340-373.

45. Stal, L. J., and Krumbien, W. E. (1985) Nitrogenase activity in the non-heterocystous cyanobacterium Oscillatoria sp. grown under alternating light-dark cycles, *Arch Microbiol 143*, 67-71.

46. Bertsova, Y. V., Bogachev, A. V., and Skulachev, V. P. (2001) Noncoupled NADH:ubiquinone oxidoreductase of Azotobacter vinelandii is required for diazotrophic growth at high oxygen concentrations, *J Bacteriol 183*, 6869-6874.

47. Seefeldt, L. C., and Arp, D. J. (1986) Purification to homogeneity of Azotobacter vinelandii hydrogenase: a nickel and iron containing alpha beta dimer, *Biochimie 68*, 25-34.

48. Johnson, D. C., Dean, D. R., Smith, A. D., and Johnson, M. K. (2005) Structure, function, and formation of biological iron-sulfur clusters, *Annu Rev Biochem 74*, 247-281.

49. Smith, A. D., Jameson, G. N., Dos Santos, P. C., Agar, J. N., Naik, S., Krebs, C., Frazzon, J., Dean, D. R., Huynh, B. H., and Johnson, M. K. (2005) NifS-mediated assembly of [4Fe-4S] clusters in the N- and C-terminal domains of the NifU scaffold protein, *Biochemistry 44*, 12955-12969.

50. Jin, Z., Heinnickel, M., Krebs, C., Shen, G., Golbeck, J. H., and Bryant, D. A. (2008) Biogenesis of iron-sulfur clusters in photosystem I: holo-NfuA from the cyanobacterium Synechococcus sp. PCC 7002 rapidly and efficiently transfers [4Fe-4S] clusters to apo-PsaC in vitro, *J Biol Chem 283*, 28426-28435.

51. Mouncey, N. J., Mitchenall, L. A., and Pau, R. N. (1995) Mutational analysis of genes of the mod locus involved in molybdenum transport, homeostasis, and processing in Azotobacter vinelandii, *J Bacteriol 177*, 5294-5302.

52. Bishop, P. E., Hawkins, M. E., and Eady, R. R. (1986) Nitrogen fixation in molybdenum-deficient continuous culture by a strain of Azotobacter vinelandii carrying a deletion of the structural genes for nitrogenase (nifHDK), *Biochem J 238*, 437-442.

53. Jacobitz, S., and Bishop, P. E. (1992) Regulation of nitrogenase-2 in Azotobacter vinelandii by ammonium, molybdenum, and vanadium, *J Bacteriol 174*, 3884-3888.

54. Premakumar, R., Jacobson, M. R., Loveless, T. M., and Bishop, P. E. (1992) Characterization of transcripts expressed from nitrogenase-3 structural genes of Azotobacter vinelandii, *Can J Microbiol 38*, 929-936.

55. Chatterjee, R., Ludden, P. W., and Shah, V. K. (1997) Characterization of VNFG, the delta subunit of the vnf-encoded apodinitrogenase from Azotobacter vinelandii. Implications for its role in the formation of functional dinitrogenase 2, *J Biol Chem 272*, 3758-3765.

56. Wolfinger, E. D., and Bishop, P. E. (1991) Nucleotide sequence and mutational analysis of the vnfENX region of Azotobacter vinelandii, *J Bacteriol 173*, 7565-7572.

57. Bageshwar, U. K., Raina, R., Choudhury, N. R., and Das, H. K. (1998) Analysis of upstream activation of the vnfH promoter of Azotobacter vinelandii, *Can J Microbiol 44*, 405-415.

58. Ruttimann-Johnson, C., Rubio, L. M., Dean, D. R., and Ludden, P. W. (2003) VnfY is required for full activity of the vanadium-containing dinitrogenase in Azotobacter vinelandii, *J Bacteriol 185*, 2383-2386.

59. Joerger, R. D., Wolfinger, E. D., and Bishop, P. E. (1991) The gene encoding dinitrogenase reductase 2 is required for expression of the second alternative nitrogenase from Azotobacter vinelandii, *J Bacteriol 173*, 4440-4446.

60. Schmehl, M., Jahn, A., Meyer zu Vilsendorf, A., Hennecke, S., Masepohl, B., Schuppler, M., Marxer, M., Oelze, J., and Klipp, W. (1993) Identification of a new class of nitrogen fixation genes in Rhodobacter capsulatus: a putative membrane complex involved in electron transport to nitrogenase, *Mol Gen Genet 241*, 602-615.

61. Kaminski, P. A., Norel, F., Desnoues, N., Kush, A., Salzano, G., and Elmerich, C. (1988) Characterization of the fixABC region of Azorhizobium caulinodans ORS571 and identification of a new nitrogen fixation gene, *Mol Gen Genet 214*, 496-502.

62. Curatti, L., Brown, C. S., Ludden, P. W., and Rubio, L. M. (2005) Genes required for rapid expression of nitrogenase activity in Azotobacter vinelandii, *Proc Natl Acad Sci U S A 102*, 6291-6296.

63. Premakumar, R., Pau, R. N., Mitchenall, L. A., Easo, M., and Bishop, P. E. (1998) Regulation of the transcriptional activators AnfA and VnfA by metals and ammonium in Azotobacter vinelandii, *FEMS Microbiol Lett 164*, 63-68.

64. Walmsley, J., Toukdarian, A., and Kennedy, C. (1994) The role of regulatory genes nifA, vnfA, anfA, nfrX, ntrC, and rpoN in expression of genes encoding the three nitrogenases of Azotobacter vinelandii, *Arch Microbiol 162*, 422-429.

65. Tabor, H., Mehler, A. H., Hayaishi, O., and White, J. (1952) Urocanic acid as an intermediate in the enzymatic conversion of histidine to glutamic and formic acids, *J Biol Chem 196*, 121-128.

66. Anbar, A. D., Duan, Y., Lyons, T. W., Arnold, G. L., Kendall, B., Creaser, R. A., Kaufman, A. J., Gordon, G. W., Scott, C., Garvin, J., and Buick, R. (2007) A

CHAPTER 6

DIFFERENTIAL ACCUMULATION OF *NIF* STRUCTURAL GENE MRNA IN *AZOTOBACTER VINELANDII*

Contributions of Authors and Co-Authors

Manuscript in Chapter 6

Author: Trinity L. Hamilton

Contributions: Conceived the study, collected and analyzed output data, and wrote the manuscript.

Co-author: Marty Jacobson

Contributions: Collected and analyzed output data, discussed results and implications and edited the manuscript.

Co-author: Marcus Ludwig

Contributions: Collected and analyzed output data, discussed results and implications and edited the manuscript.

Co-author: Eric S. Boyd

Contributions: Discussed the results and implications and edited the manuscript at all stages.

Co-author: Donald A. Bryant

Contributions: Discussed the results and implications and edited the manuscript at all stages.

Co-author: Dennis R. Dean

Contributions: Discussed the results and implications and edited the manuscript at all stages.

Co-author: John W. Peters

Contributions: Obtained funding, discussed the results and implications and edited the manuscript at all stages.

Manuscript Information Page

Authors: Trinity L. Hamilton, Marty Jacobson, Marcus Ludwig, Eric S. Boyd, Donald A. Bryant, Dennis R. Dean, and John W. Peters

Journal: Journal of Bacteriology

Status of the Manuscript:

___Prepared for submission to a peer-reviewed journal
___Officially submitted to a peer-reviewed journal
___Accepted by a peer-reviewed journal
_x_Published in a peer-reviewed journal

Publisher: American Society for Microbiology

First published July 2011, doi: 10.1128/JB.05100-11

CHAPTER 6

DIFFERENTIAL ACCUMULATION OF *NIF* STRUCTURAL GENE MRNA IN *AZOTOBACTER VINELANDII*

Published: *Journal of Bacteriology*, **2011**, doi: 10.1128/JB.05100-11

Trinity L. Hamilton[1], Marty Jacobson[3], Marcus Ludwig[2], Eric S. Boyd[1], Donald A. Bryant[1,2], Dennis R. Dean[3*], and John W. Peters[1*]

[1]*Department of Chemistry and Biochemistry and the Astrobiology Biogeocatalysis Research Center, Montana State University, Bozeman, Montana 59717*
[2]*Department of Biochemistry and Molecular Biology, The Pennsylvania State University,*
[3]*Department of Biochemistry, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061*

Abstract

Northern analysis was employed to investigate mRNA produced by mutant strains of *Azotobacter vinelandii* with defined deletions in the *nif* structural genes and in the intergenic noncoding regions. The results indicate that intergenic RNA secondary structures effect the differential accumulation of transcripts supporting the high Fe protein to MoFe protein ratio required for optimal diazotrophic growth.

Differential Accumulation Of *nif* Structural Gene mRNA In *Azotobacter Vinelandii*

Biological nitrogen fixation occurs by the activity of nitrogenase, which exists as a complex metalloenzyme composed of two easily separable components. The Fe protein component (encoded by *nifH*) serves as the obligate electron donor to the MoFe protein component (encoded by *nifDK*), which contains the active site for dinitrogen reduction (*1*). In addition to the structural proteins, the nitrogenase enzyme requires extensive biosynthetic machinery consisting of enzymes, scaffolds and carrier proteins to assemble the metalloclusters required for catalysis (*2, 3*). *nif*-encoded genes are located in two

clusters in the model diazotroph, *A. vinelandii*. The major *nif* cluster encodes the structural genes and most of the biosynthetic machinery which are organized in several contiguous operons (*4*), and a minor *nif* cluster contains the *nif* regulatory elements and the remaining, necessary biosynthetic genes (*5*). The major *nif* cluster in *A. vinelandii* was sequenced over three decades ago and has been the subject of a number of gene deletion and mutation studies, which laid the groundwork for our current understanding of the operon structure and regulation of *nif (4)*. The genes encoding the nitrogenase structural proteins, *nifH*, *nifD* and *nifK*, are located in the major *nif* operon and are co-transcribed from a single promoter, the *nifH* promoter, along with *nifT*, *nifY*, *orf1*, and *lrv* (*4*). The *nifH* promoter is efficient and effectively regulated driving the rapid and abundant expression of nitrogenase in the absence of fixed N and conversely, an abrupt decline in nitrogenase component transcription in the presence of fixed N (*6*).

In vitro characterization of nitrogenase from a variety of microbial sources indicates that highest nitrogenase activities are observed at high Fe protein to MoFe protein ratios. In line with these observations, northern blot hybridization analyses have indicated differential expression of the *nif* structural proteins (*7*), and immunoblotting has shown that the Fe protein occurs in significant excess compared to MoFe protein levels in cultures grown under diazotrophic conditions (*7-10*). Moreover, global transcriptional analyses indicate *nifH* mRNA accumulates at ~ 3-fold higher levels than *nifD* and *nifK* messages in *A. vinelandii* grown diazotrophically (*11*). Because the genes encoding the nitrogenase Fe protein and MoFe protein are co-transcribed in *A. vinelandii*, a mechanism must exist to control differential *nif* structural gene transcript abundance.

In order to assess possible mechanisms resulting in the differential accumulation of *nif* structural gene mRNA and nitrogenase component proteins, northern blot hybridization analysis was performed probing mRNAs encoding *nifH*. Wild-type *A. vinelandii* and deletion strains were grown in fixed-N replete, modified Burk medium (*12*), and cells were de-repressed to induce expression of nitrogenase as previously described (*6*). Deletion strains were constructed as previously described (*4, 13*). Samples for northern blot hybridization were collected at 10-min intervals for 120 min. RNA extraction, glyoxylation and electrophoresis were performed as described (*14, 15*). Transfer of RNA onto GeneScreen^{TM} hybridization membranes and subsequent hybridization were performed according to the manufacturers instructions (Dupont). Approximately 10 μg of RNA from each sample was analyzed by northern blot hybridization using a *nifH*-specific, ^{32}P-labelled probe pMJH5 purified from *E. coli* as previously described (*7*). The results revealed differential accumulation of major transcripts that could correspond to *nifH*, *nifHD, nifHDK,* and a very minor transcript corresponding to the length of the entire operon, *nifH*, *nifD, nifK, nifT*, *nifY*, *orf1*, and *lrv* (Fig. 6.1A). The identity of the three most abundant transcripts was established in two ways. First, when a *nifD*-specific hybridization probe was used for northern blot analysis, only the bands corresponding to putative *nifHD* and *nifHDK* transcripts were detected (data not shown).

In a second series of experiments, strains having defined deletions within the structural gene region were analyzed by northern blot analysis using a *nifH* probe (Fig. 6.1B). These results showed that a deletion within *nifD* had no effect on the size of the accumulated *nifH* transcripts but resulted in accumulation in smaller transcripts assigned to *nifHD* and *nifHDK* (DJ100, Fig. 6.1B). Similarly, a deletion in *nifK* had no effect on

the size of the transcripts corresponding to *nifH* or *nifHD* but resulted in the accumulation of a smaller transcript assigned to *nifHDK* (DJ13, Fig. 6.1B). Strains that had deletions that spanned portions of *nifH* and *nifD* (DJ46, Fig. 6.1B) or *nifD* and *nifK* (DJ33, Fig. 6.1B) resulted in the accumulation of only two major transcripts, which suggested that elements leading to differential accumulation are likely to be located within the intergenic regions. A strain having a deletion of *nifH* that encompassed the region corresponding to the *nifH*-specific probe showed no *nifH* hybridizing transcripts (DJ54, Fig. 6.1B). These results are consistent with the aforementioned results showing that the major *nif* operon produces three major transcripts that correspond to *nifH*, *nifHD*, and *nifHDK*.

It is interesting that during the transition to nitrogen fixing conditions, the relative abundance of the three *nifH* specific transcripts changed noticeably. For example, inspection of the time course shown in Fig. 1A revealed that, in the early stage of *nif*-derepression, the major accumulating mRNA corresponds to *nifHDK*. In contrast, as the cells enter steady-state, nitrogen fixation conditions (~90 min following derepression), the relative amounts of shorter transcripts encoding solely *nifH* increased in abundance relative to the longer transcripts. The elevated transcript levels for *nifH*, compared to *nifDK* expression under steady state nitrogen fixing conditions is in line with establishing and maintaining the high Fe protein to MoFe protein ratios required for optimal nitrogenase catalytic activity. It is not clear why there is an apparent increased capacity for expression of *nifD* relative to *nifK*, because these genes encode subunits of the MoFe protein, which are required in equimolar amounts. However, it could be a mechanism to counteract potential differences in translation efficiencies for these transcripts or in the stabilities of nitrogenase protein subunits.

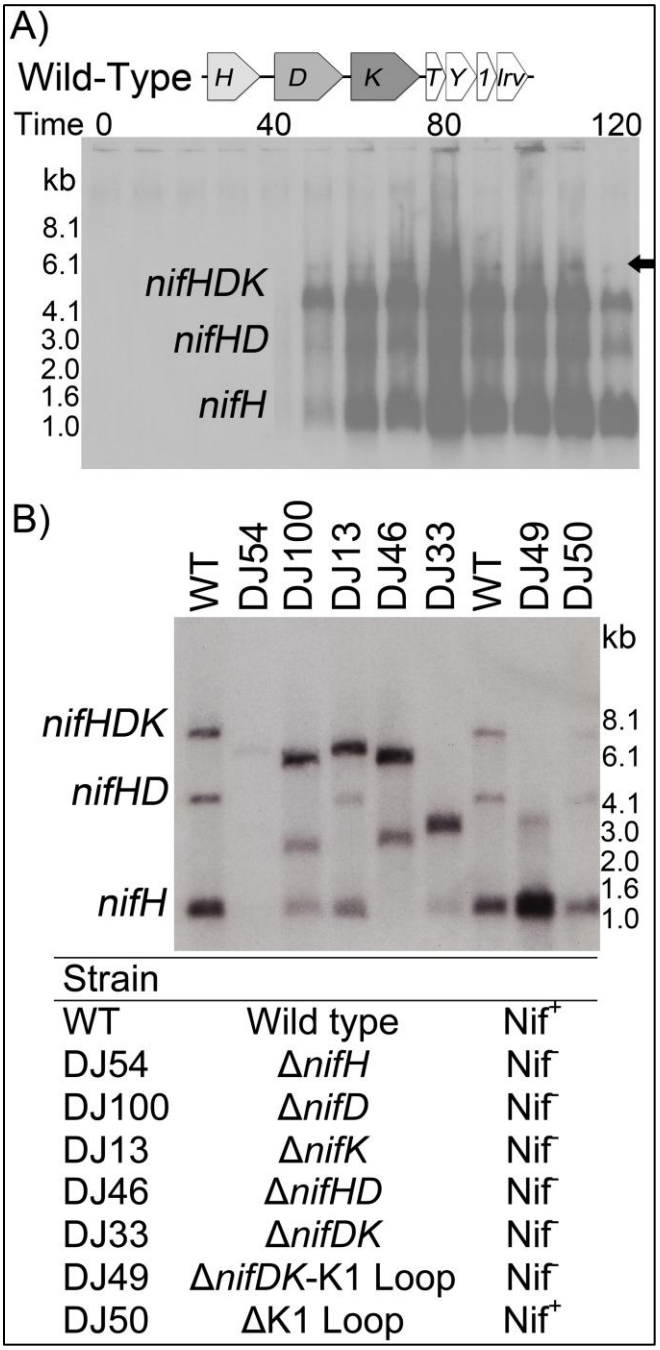


Figure 6.1. (A) Northern blot hybridization analysis of total RNA isolated from wild-type *A. vinelandii*. Cells were de-repressed in N-free media and total RNA was extracted at 10-min increments from 0 to 120 min. Black arrow indicates the minor band resulting from the entire structural gene operon. (B) Northern blot hyrbidization analysis of total RNA isolated from *A. vinelandii* strains with deletions spanning regions of the structural genes indicated. Cells were de-repressed in N-free media and total RNA was extracted after 60 min. Nif^+ were capable of diazotrophic growth, Nif^- were not.

A closer analysis of the nucleotide sequences in the regions that separate the coding regions of the *nif* structural gene operon revealed potential secondary structures between *nifH* and *nifD*, between *nifD* and *nifK*, and downstream from *nifK* ((*16, 17*); Fig. 6.2A). Free energy (ΔG) values of -29.0 and -26.9 kcal/mol were calculated for the structures between *nifH* and *nifD* and *nifD* and *nifK*, respectively (Fig. 6.2A). Additional structures predicted just downstream of *nifK* had free energy values of -29.0 and -44.0 kcal/mol. Transcript profiling of *A. vinelandii* under Mo-dependent, nitrogen fixing conditions (*11*) indicates a specific decrease in transcript sequences corresponding to a very small region that spans the putative mRNA secondary structure (Fig. 6.2A). Similar decreases in transcript abundance were observed for the regions corresponding to the predicted structures between *nifD* and *nifK* and downstream of *nifK* (data not shown). Conventionally, RNA structures have been probed individually but the advent of whole transcriptome sequencing has precipitated the need for high-throughput RNA structure probing to complement transcriptional profiling (*18, 19*). However, mapping the end of the most abundant message, corresponding to *nifH* clearly indicates that transcript levels declined upstream of the first structure observed in the operon located between *nifH* and *nifD* (Fig 6.2A). An inability to produce any detectable cDNA spanning the proposed secondary structure located within the *nifH – nifD* intergenic region provides strong evidence for the formation of a secondary structure within this region. This is further supported by the failure to detect cDNA sequence reads within the regions that define the putative secondary structures in the *nifD – nifK* integenic region and downstream of *nifK* (data not shown). To examine the role of the intergenic regions in the differential accumulation of *nif* structural gene operon directly, a strain was constructed (DJ81) that

had an 87 base-pair deletion spanning the proposed secondary structure located in the *nifH-nifD* intergenic region. Strain DJ81 was capable of diazotrophic growth at rates comparable to wild-type under typical laboratory culture conditions (*4*) (data not shown). Northern blot hybridization analysis of this strain revealed the accumulation of only two major *nifH* transcripts corresponding the *nifHD* and *nifHDK* and a very minor transcript that could correspond to *nifH, nifD, nifK, nifT, nifY, orf1,* and *lrv* (Fig. 6.2B). In this strain, no *nifH*-only transcripts were detected. Similar to wild-type (Fig. 6.1A), the first major accumulating mRNA during *nif*-derepression of DJ81 corresponds to *nifHDK*, with the shorter *nifHD* transcript accumulating later. This result provides further evidence that the secondary structures act as processing and/or mRNA stabilizing sites that increase the lifetime of specific segments of the primary transcripts.

In summary, the results presented here confirm that the genes within the *nif* structural gene operon are co-transcribed from a single promoter and clearly demonstrate the role of intergenic RNA secondary structures in differential accumulation of *nif* structural gene mRNAs. The intergenic secondary structures could function as premature termination sites and/or processing and stabilization. The changes in the relative abundances of *nifH* specific transcripts that are observed to occur over the course of depression (Fig. 6.1A and Fig. 6.2B) seems to favor the latter or imply the involvement of *trans*-acting elements in controlling the ratios of these transcripts for optimal diazotrophic growth. In this regard, the differential accumulation of mRNA represents one mechanism for the production of the appropriate relative abundances of the nitrogenase components for optimal catalytic function.

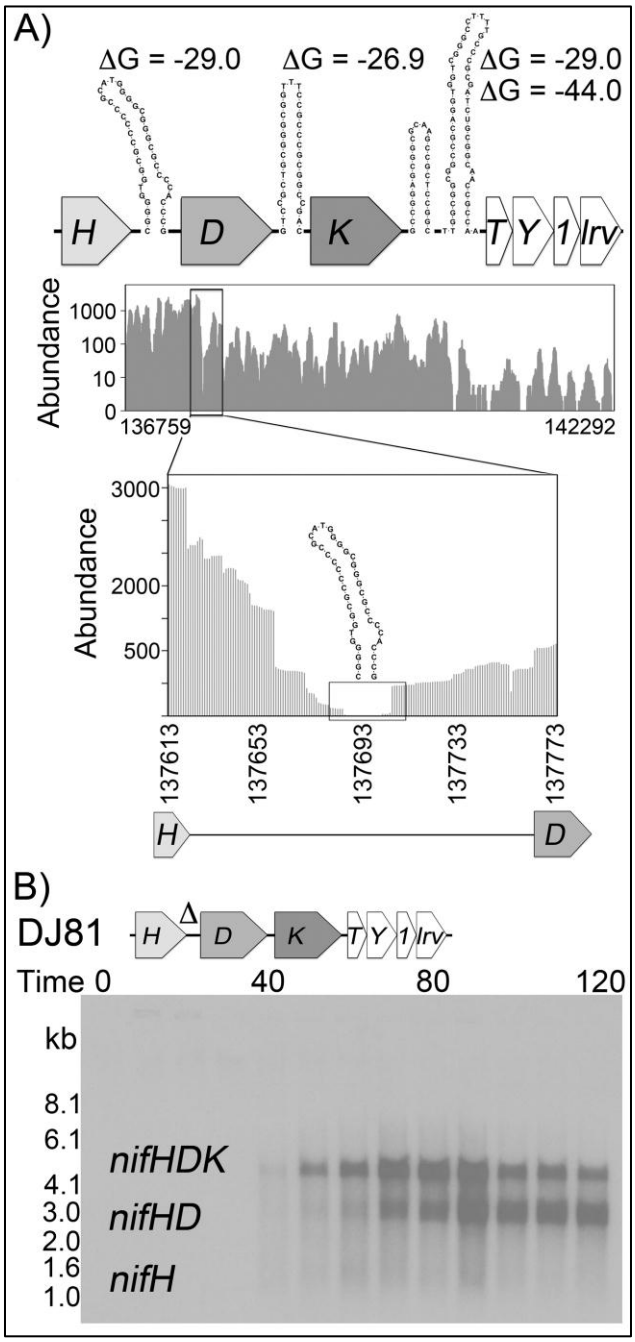


Figure 6.2. (A) Identification of RNA secondary structures between the *nif* structural genes. ΔG values are given in kcal/mol. Transcript abundance of the intergenic region between *nifH* and *nifD* plotted against the location of these genes in the genome. The box indicates the location of the RNA secondary structure. For detailed methods of the transcriptional profiling, see reference (4). (B) Northern blot hybridization analysis of total RNA isolated from mutant strain DJ81. Cells were de-repressed in N-free media and total RNA was extracted at 10-min increments from 0 to 120 min.

This mechanism may also be important in compensating for differences in translation efficiencies of the *nif* structural genes or protein subunit stabilities.

Acknowledgements

This work was supported by the NASA Astrobiology Institute grant NNA08C-N85A to J.W.P. and NASA Astrobiology: Exobiology and Evolutionary Biology, Award NNX09AM87G (D.A.B.) and NSF MCB-071770 (D.R.D). T.L.H. was supported by an NSF-Integrated Graduate Educational Research and Training fellowship grant and E.S.B. was supported by a fellowship from the NASA Astrobiology Institute Postdoctoral Program.

Literature Cited

1. Peters, J. W., and Szilagyi, R. K. (2006) Exploring new frontiers of nitrogenase structure and mechanism, *Curr Opin Chem Biol 10*, 101-108.

2. Dos Santos, P. C., and Dean, D. R. (2011) Coordination and fine-tuning of nitrogen fixation in Azotobacter vinelandii, *Molecular Microbiology*.

3. Rubio, L. M., and Ludden, P. W. (2008) Biosynthesis of the iron-molybdenum cofactor of nitrogenase, *Annu Rev Microbiol 62*, 93-111.

4. Jacobson, M. R., Brigle, K. E., Bennett, L. T., Setterquist, R. A., Wilson, M. S., Cash, V. L., Beynon, J., Newton, W. E., and Dean, D. R. (1989) Physical and genetic map of the major nif gene cluster from Azotobacter vinelandii, *J Bacteriol 171*, 1017-1027.

5. Joerger, R. D., and Bishop, P. E. (1988) Nucleotide sequence and genetic analysis of the nifB-nifQ region from Azotobacter vinelandii, *J Bacteriol 170*, 1475-1487.

6. Shah, V. K., Davis, L. C., and Brill, W. J. (1972) Nitrogenase. I. Repression and derepression of the iron-molybdenum and iron proteins of nitrogenase in Azotobacter vinelandii, *Biochim Biophys Acta 256*, 498-511.

7. Jacobson, M. R., Premakumar, R., and Bishop, P. E. (1986) Transcriptional regulation of nitrogen fixation by molybdenum in Azotobacter vinelandii, *J Bacteriol 167*, 480-486.

8. Dingler, C., Kuhla, J., Wassink, H., and Oelze, J. (1988) Levels and activities of nitrogenase proteins in Azotobacter vinelandii grown at different dissolved oxygen concentrations, *J Bacteriol 170*, 2148-2152.

9. Robinson, A. C., Burgess, B. K., and Dean, D. R. (1986) Activity, reconstitution, and accumulation of nitrogenase components in Azotobacter vinelandii mutant strains containing defined deletions within the nitrogenase structural gene cluster, *J Bacteriol 166*, 180-186.

10. Robinson, A. C., Dean, D. R., and Burgess, B. K. (1987) Iron-molybdenum cofactor biosynthesis in Azotobacter vinelandii requires the iron protein of nitrogenase, *J Biol Chem 262*, 14327-14332.

11. Hamilton, T. L., Ludwig, M., Dixon, R., Boyd, E. S., Dos Santos, P. C., Setubal, J. C., Bryant, D. A., Dean, D. R., and Peters, J. W. (2011) Transcriptional profiling of nitrogen fixation in Azotobacter vinelandii, *J Bacteriol 193*, 4477-4486.

12. Strandberg, G. W., and Wilson, P. W. (1968) Formation of the nitrogen-fixing enzyme system in Azotobacter vinelandii, *Can J Microbiol 14*, 25-31.

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

14. Krol, A. J., Hontelez, J. G., Roozendaal, B., and van Kammen, A. (1982) On the operon structure of the nitrogenase genes of Rhizobium leguminosarum and Azotobacter vinelandii, *Nucleic Acids Res 10*, 4147-4157.

15. McMaster, G. K., and Carmichael, G. G. (1977) Analysis of single- and double-stranded nucleic acids on polyacrylamide and agarose gels by using glyoxal and acridine orange, *Proc Natl Acad Sci U S A 74*, 4835-4838.

16. Hofacker, I. L. (2003) Vienna RNA secondary structure server, *Nucleic Acids Res 31*, 3429-3431.

17. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction, *Nucleic Acids Res 31*, 3406-3415.

18. Kertesz, M., Wan, Y., Mazor, E., Rinn, J. L., Nutter, R. C., Chang, H. Y., and Segal, E. (2003) Genome-wide measurement of RNA secondary structure in yeast, *Nature 467*, 103-107.

19. Underwood, J. G., Uzilov, A. V., Katzman, S., Onodera, C. S., Mainzer, J. E., Mathews, D. H., Lowe, T. M., Salama, S. R., and Haussler, D. (2010) FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing, *Nat Methods 7*, 995-1001.

CHAPTER 7

CONCLUDING REMARKS

Biological nitrogen fixation had a profound impact on the evolution of life on Earth. Because fixed nitrogen is essential to all life, the origin and evolution of biological the nitrogenase enzyme has been of extreme interest. Geochemical and isotopic analyses have resulted in several hypotheses regarding the emergence of biological nitrogen fixation. These analyses offer important clues into the type of environment where fixed nitrogen limitation provided the impetus for emergence of the nitrogenase enzyme. Combining evidence for early Earth conditions favoring the emergence of biological nitrogen fixation and the vast amount of biochemical and genetic characterization of the nitrogenase enzyme itself, this work has yielded key insight into the origin of Mo-nitrogenase, the environmental constraints on the distribution and diversity of diazotrophy and the cross-talk necessary for assembly and activation of the alternative nitrogenases.

Owing to the complexity of the enzyme as well as the biosynthetic machinery, the component proteins necessary for nitrogenase maturation and activity are very well-conserved and amenable to evolutionary analyses. In chapter 2, phylogenetic analyses of concatenated structural proteins H, D, and K from *anf*, *vnf*, and *nif* operons as well as HDK homologs found in organisms not known to fix nitrogen resulted in well-supported lineages that correspond to active site metal content. In this analyses, *nif*-encoded HDK proteins formed two lineages, one at the base of the tree comprised solely of hydrogenotrophic methanogens and the other comprised of all other characterized *nif*-

encoded HDK, all of which are more recently evolved. The alternative *vnf*-encoded HDK and *anf*-encoded HDK proteins were nested within two Nif sublineages, indicating that Vnf and Anf are derived from Nif. Furthermore, AnfHDK are nested within VnfHDK, indicating Anf was derived from Vnf. These results provide strong evidence that Mo-nitrogenase is the ancestor of the alternative nitrogenases, a finding that contrasts what has been previously proposed based on metal availability under early Earth conditions. Also in Chapter 2, the high degree of sequence conservation of these proteins was used to elucidate key phylogenetic events such as gene fusions and gene duplications of the biosynthetic and structural genetic machinery that yield insight into the origin and evolution of the nitrogenase enzyme. Phylogenetic reconstruction in which NifN is nested among NifK and NifE among NifD provides evidence that *nifDK* duplicated to give rise to *nifEN*. In addition, NifEN in Bacteria and Archaea were not reciprocally monophyletic. The phylogenetic branching order NifDKEN indicates that the duplication of *nifDK* to *nifEN* most likely occurred in an ancestor of the hydrogenotrophic methanogens and was acquired in bacteria via a lateral gene transfer event. Examining substitution rates of the proteins involved in (bacteriol)chlorophyll biosynthesis, BchNBYZ, compared to NifDKEN indicates that the Mo-nitrogenase emerged 1.5 to 2.2 Ga, after the origin of oxygenic photosynthesis. Together, these data provide strong evidence that Mo-nitrogenase was not present in LUCA and is a relatively recent evolutionary innovation.

The highly conserved nature of the nitrogenase enzyme has enabled the design and implementation of a PCR-based approach to examine the environmental parameters that underpin the distribution and diversity of diazotrophs in natural environments.

Chapter 3 details the development of this approach in Yellowstone National Park which provides abundant geochemical gradients, enabling a better understanding of the environmental parameters that drive the evolution of biological nitrogen fixation. These data indicate the phylogenetic diversity of putative diazotrophs is constrained at the broadest level by geographical barriers, such as those imposed by the geographically discontinuous nature of geothermal features in YNP. Environmental factors such as salinity impose niche conservatism on *nifH* assemblages, which has occurred over geological time scales.

The distribution of putative diazotrophs in the Yellowstone geothermal complex was widespread and was not constrained by temperature or pH. The recovery of *nifH* amplicons from high temperature low pH springs (detailed in chapter 3) was not expected. In chapter 4, the activity of diazotrophs from some of these high temperature, low pH environments was characterized. Partial enrichment, gene abundance data and temperature- and pH-responsive activity assays indicate diazotrophic populations have adapted to local environmental conditions.

In chapter 5, transcriptional profiling experiments were carried out in *A. vinelandii*, a model diazotroph, to determine the response of this organism to diazotrophic conditions. The recent completion of the genome of *A. vinelandii* and the advent of full transcriptome sequencing enabled a comprehensive analysis of transcript abundance as a result of metal availability in the absence of fixed nitogen. This study revealed the suite of genes necessary for diazotrophic growth in the absence of Mo, providing key insight into the crosstalk between nitrogenase biosynthetic machinery as well as indicating the effects of biological nitrogen fixation on central metabolism in this organism.

Chapter 6 details the differential accumulation of *nif* structural gene mRNA, combining transcript abundance data with northern blot hybridization analyses. This study identified intergenic RNA secondary structures that effect the transcription of *nif* structural genes, indicating these secondary structures lead to the high Fe protein to MoFe protein ratio necessary for optimal diazotrophic growth.

Collectively, the results presented in this dissertation provide key insight into the emergence and evolution of biological nitrogen fixation. The combination of phylogenetic analyses and with decades of biochemical and genetic characterization enabled a comprehensive analysis of the evolution of nitrogenase, the occurrence of diazotrophy in geochemically diverse environments, and the intricate cross-talk of biosynthetic machinery necessary for metal-limited diazotrophic growth in *A. vinelandii*. The phylogenetic analyses presented here provide significant evidence that Mo-nitrogenase is more recently evolved than has been previously proposed. In addition, these data strongly suggest the alternative forms of nitrogenase are not the ancestral forms and that these arose from Nif, again contrary to what has been proposed based on geochemical evidence. Examination of nitrogen fixation in high temperature acidic springs in chapter 4 is the highest temperature diazotrophy observed to date in terrestrial ecosystems and also represents the lowest pH under which the process has been observed. Full transcriptome sequencing of *A. vinelandii* indicated the alternative forms of nitrogenase, Anf and Vnf, rely on *nif*-encoded biosynthetic machinery for diazotrophic growth in the absence of Mo. These analyses deomonstrated that diazotrophy in *A. vinelandii* is regulated by metal availability and further supported that Nif is preferred over the alternative forms. These data support the phylogenetic analyses which cannote

that Anf and Vnf arose from Nif because the alternative systems do not encode the full complement of biosynthetic machinery. The work presented herein provides significant insight into the origin, evolution, and ecology of biological nitrogen fixation and demonstrate the utility of a combined approach of phylogenetics, structural biology, biochemical characterization, genetics, geochemistry, and ecology to examine the emergence and evolution of enzymes containing complex metallocofactors.

APPENDICES

APPENDIX A:

SUPPORTING INFORMATION FOR CHAPTER 2

Chapter 2: Examining the Origin and Evolution of Biological Nitrogen Fixation

Published: Geobiology, **2011**, *9*, 221 – 232. Frontier in Microbiology, **2011**, doi: 10.3389/fmicb.2011.00205

Further Supporting Materials are available free of charge via the Internet at http://onlinelibrary.wiley.com/doi/10.1111/j.1472-4669.2011.00278.x/suppinfo and http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3207485/. This material contains additional methods and results sections as well as tables of the accession number of the genes included in the phylogenetic analyses and additional figures.

Methods

Phylogenetic Analysis

Representative homologs of Anf/Vnf/NifHDK (Table S1), as well as representative homologs of uncharacterized HDK, were compiled as previously described (*1*, *2*). Individual H, D, and K homologs were aligned using CLUSTALX (version 2.0.8) specifying the Gonnet 250 protein substitution matrix and default gap extension and opening penalties (Larkin et al., 2007). ChlLNB/BchLNB from *Anabaena variabilis* ATCC 29413 and *Chlorobium limicola* DSM 245, which are paralogs of Anf/Vnf/NifHDK, respectively, were usedas the outgroups. Each alignment was scrutinized and manually aligned using known catalytic residues as previously described (*1*). The H, D, and K alignment blocks were concatenated and the concatenated alignment block was subjected to evolutionary model prediction using ProtTest (version 2.0) (*3*). ProtTest identified the Whelson and Goldman (WAG) evolutionary model with gamma distributed rate variation with a proportion of invariable sites (I+G) as the best fit model for the data. The phylogeny of each protein sequence was evaluated using MrBayes (ver. 3.1.2) (*4*) employing the WAG+I+G model. Tree topologies were sampled every 500 generations over 450,000 generations (after a burnin of 50,000) at likelihood stationarity and after convergence of two separate Markov chain Monte Carlo runs (average standard deviation of split frequencies <0.05). A consensus phylogenetic tree was projected from 1800 trees using FigTree (version 1.2.2) (http://tree.bio.ed.ac.uk/UH). PhyML (ver. 3.0) (*5*) was also used to infer the phylogenetic position of concatenated HDK sequences with the WAG+I+G model

specified and 100 bootstrap replicates. Matrices describing the phylogenetic dissimilarity of concatenated HDK homologs, inferred using both MrBayes and PhyML, were generated using Phylocom (version 4.0.1) (*6*).

Structural Studies

The structures of the representative H, D, and K homologs selected for phylogenetic analysis were inferred using the CPH homology server for protein homology modeling (*7*) using the NifH (*8*), NifD (*9-11*), and NifK (*9-11*) crystal structures from *Azotobacter vinelandii* AvOP and *Klebsiella pneumoniae* as references. The inferred structures for each H, D, and K homolog were imported into PyMol (ver. 1.4) (http://www.pymol.org/). The root-mean-square-deviations (RMSD) in the C^{α} positions were calculated for each individual inferred H, D, and K homolog structure in relation to the other inferred H, D, or K homolog structures, resulting in a pairwise matrix describing the structural RMSD (e.g., structural dissimilarity) for H, D, and K homologs. RMSDs generated for H, D, and K homologs, were normalized to compensate for differing HDK protein lengths, and the normalized H, D, and K matrices were then were averaged to produce an HDK RMSD matrix for use in statistical analyses. PyMol was also used to generate images of sequence conservation in the active site cavity of Anf/Vnf/NifDK homologs, and to infer the structure of the ancestral B/D-N/K protein sequence.

Statistical Analyses

Mantel regressions were performed using XL Stat (ver. 2009.5.01). Ten thousand permutations employing two-tailed t-tests were used to determine the strength and significance of the relationships between dissimilarity matrices, respectively.

Vnf/Anf/NifDKEN and NifB Phylogenetic Analyses

BLASTp was used to compile all D, K, E, N, and B deduced amino acid sequences from genomic sequences using the DOE-IMG and the NCBI Genome Blast servers (Supp. Table 1). Those which contained a full complement of DKENB were retained (Vnf and Nif operons) or DKB (Anf operons). Each individual locus was aligned using ClustalX (version 2.0.8) with the Gonnet 250 protein matrix and default gap extension and opening penalties (*12*). Each alignment was scrutinized and manually aligned using known catalyztic residues when available (*13*, *14*). ProtTest (version 2.0) (*3*) was used to select WAG+I+G+F as the best-fit protein evolutionary model for individual Vnf/Anf/NifDKEN and NifB proteins and to determine the amino acid frequency (specified as 'F' in the model). The phylogeny of each locus was evaluated using MrBayes (version 3.1.2) (*15*, *16*) using the WAG evolutionary model with fixed (F) amino acid frequencies and gamma-distributed rate variation with a proportion of invariable sites (I+G). Tree topologies were sampled at likelihood stationarity (average standard deviation of split frequencies less than 0.05) during 2 separate runs every 500 generations by MrBayes over 2.0e6 generations (burnin=1.0e6). Taxa with D, K, E, and N loci that represented each of the major lineages were empirically selected (Supp. Table 2) and these loci were subjected to evolutionary analysis as described above (See Supp.

Fig. 1 legend). A composite DKEN phylogram was constructed using the WAG evolutionary model with fixed amino acid frequencies and gamma-distributed rate variation with a proportion of invariable sites (I+G). Tree topologies were sampled at likelihood stationarity during 2 separate runs every 500 generations by MrBayes over 2.0e6 generations (burnin=1.0e6). The unrooted phylogram with collapsed clades (Fig. 1) was projected using TreeView (version 1.6.6) (*17*). The radial phylogram (Supp. Fig. 1) was projected with FigTree (version 1.2.2) (HUhttp://tree.bio.ed.ac.uk/UH). The orientation of *nif* regulons from *Bradyrhizobium japonicum* and *Burkholderia phymatum* have been reversed to conserve space. The genomic context of nitrogenase regulons was determined manually using the IMG Gene Neighborhood viewer.

Vnf/Anf/NifDKEN and BchNBYZ Phylogenetic Analyses

BLASTp was used to compile all BchNBYZ deduced amino acid sequences from genomic sequences using the DOE-IMG and the NCBI Genome Blast servers (Supp. Table 3). To simplify the analysis, bacteriochlorophyll biosynthesis protein sequences from eukarya were not considered in this analysis since previous studies indicate phototrophic eukarya to be derived from phototrophic bacteria (*18*). Sequence alignment, evolutionary model selection, and phylogenetic analyses were performed as described above for each individual locus. Taxa representing the primary protein lineages in each individual BchNBYZ locus and representative Vnf/Anf/NifDKEN loci were used to construct a composite Vnf/Anf/NifDKEN/BchNBYZ phylogram using the WAG evolutionary model with fixed (F) amino acid frequencies and gamma-distributed rate variation with a proportion of invariable sites (I+G) (WAG+I+G+F as specified by

ProtTest). The radial phylogram (Supp. Fig. 3A) was projected with FigTree. Tree topologies were sampled at likelihood stationarity during 2 separate runs every 500 generations by MrBayes over 2.0e6 generations (burnin=1.0e6). All alignment files and tree files can be obtained through email contact with the author (eboyd@montana.edu).

Evolutionary Rates Analysis

Estimates of rates of substitution were converted to absolute ages for individual crown clades using penalized-likelihood (PL), rate constancy, and non-parametric approaches (*19*) implemented in version 1.71 of the program r8s (*19*). Each method yielded similar estimated age mean and variances. The results of PL are presented with the root age was fixed to a unit-less age of 1.0 such that it could be scaled by actual estimates of the age of the common ancestor of the nitrogenase and photosynthetic loci (*e.g.*, 3.5 to 4.0 Ga). The rate-smoothed chronogram (Supp. Fig. 3B) was projected with FigTree.

Results

We inferred the protein structures of HDK homologs from *anf*, *vnf*, *nif*, and uncharacterized operons using homology modeling based on the NifHDK from *A. vinelandii*. Pairwise comparisons of the inferred protein structures enabled the generation of a matrix that describes their structural dissimilarity. A regression of this matrix and a matrix describing the phylogenetic dissimilarity of the concatenated HDK proteins revealed a significant and positive relationship (Mantel $R^2 = 0.23$, $p < 0.01$). This indicates that the structure of nitrogenase has evolved significantly through time.

However, the slope of the linear regression indicated that a ~2 unit increase in phylogenetic dissimilarity resulted in only a ~1 unit increase in structural dissimilarity, suggesting that deviation in inferred tertiary structure is constrained to a greater extent than that observed in the primary sequence. Therefore, we examined the conservation of residues that line the active site pocket in Anf/Vnf/NifD to determine if the active site cavity has evolved to accommodate the varying cluster compositions associated with Anf, Vnf, Nif, and to potentially uncover evidence that could provide insight into the composition of the active site cofactor in uncharacterized nitrogenase homologs. This analysis revealed a number of residues that line the active site cavity that are conserved among Anf, Vnf, Nif, and uncharacterized nitrogenase including the two covalent ligands to the FeFe-, FeV-, and FeMo-cofactors Cys-275 and His-442 (numbered according to *A. vinelandii* NifD). In addition to ligand, a number of other residues in the binding cavity are conserved, including Val-70, Gln-191, His-195, and Phe-391. Together, these results suggest that once the active site cavity was evolved, it was maintained despite differences in the metal composition of the cofactor.

We next compared the conservation in these residues with NflD, a nitrogenase homolog that has been hypothesized to be involved in F_{430} biosynthesis (*20*), and BchN, a homolog involved in bacteriochlorophyll biosynthesis (*21*). A previous phylogenetic reconstruction of Anf/Vnf/NifHD, BchLN, and NflHD revealed three lineages, with NflHD proteins forming a lineage that bisected the a lineage comprising Anf/Vnf/NifHD and a lineage comprising BchLN (*22*), suggesting that that NflHD is ancestral to Anf/Vnf/NifHD and BchLB (*20*). Intriguingly, the active site cavity of NflD proteins shared little sequence conservation with that of Anf/Vnf/NifHD and BchLN. Likewise,

the FeMo-co ligands in Anf/Vnf/NifD are not conserved in BchN sequences, although the crystal structure of BchNB reveals an open cavity for the binding of protochlorophyllide that is in the same position as to where FeMo-co is bound in nitrogenase. Thus, the two derived states (Anf/Vnf/NifD and BchN) have maintained similar structural architecture, but appear to have fine-tuned cavity residues to bind target substrates. To examine whether the open cavity architecture was a structural property of the ancestor of Anf/Vnf/NifD and BchN, we generated homology models of NflD from *Methanocalodococcus jannaschii* DSM 2661 (Accession no. NP_248427) threaded on *A.vinelandii* NifD and BchN from *Rhodobacter capsulatus* SB 1003 (Accession no.YP_003576837). The inferred structures reveal a cavity in NflD that is similar to that of BchN, the size of which would be capable of binding F_{430}. These analyses suggest that nitrogenase evolved from an ancestral protein that exhibited an open cavity and adapted that cavity to covalently ligate the FeMo-co necessary for N_2 reduction.

Anf/Vnf/NifDK Branching Order

NifD proteins from Mo-dependent methanogens form a very early branching group that along with AnfD and VnfD, form a lineage that branches distinctly from bacterial and *Methanosarcinales* NifD ("*Ms*" versus "*Mb*/*Mc*" lineages labeled in Fig. 1C).

Assuming equal rates of evolution, evidence that VnfD diversified before AnfD is seen in the deeper crown clade of VnfD compared to AnfD, which is mirrored in the paralogous earliest K/N branching lineage, where the well-supported AnfK clade is nested within a more diverse VnfK group. This observation may indicate that AnfDK results from an in-

tandem duplication of VnfDK. This interpretation is bolstered by the low posterior probability supporting the Vnf/AnfK lineages, which is evidence that the three sublineages could just as well form a grade at the base of the K/N lineage (Fig. 1C), where VnfK is resolved paraphyletic in contrast to the monophyletic AnfK. Vnf/AnfK branch early when compared to methanogen NifK, suggesting that alternative nitrogenase emerged prior to Mo-dependent nitrogenase. Similarly, the VnfEN lineage (see discussion below) which contains two constituent and divergent sub-lineages branches near the root of the DKEN phylogram (Fig. 1C), suggesting these proteins to have diverged very early with respect to Vnf/Anf/NifKEN proteins. Importantly, while the phylogenetic branching order of Anf/Vnf/NifKEN proteins supports an origin for alternative nitrogenase that predates Mo-nitrogenase, the branching order of Anf/Vnf/NifD proteins neither supports nor refutes this interpretation.

VnfEN

The proteins comprising the VnfEN lineage are ~10-20% identical to those comprising the Anf/VnfDK and NifDKEN lineages, indicating that they have a shared ancestry. Importantly, while the VnfEN lineage is annotated as *vnfE* and *vnfN*, respectively, there has not been a single biochemical study targeting this group of proteins that would indicate that they perform an analogous role as NifEN [*Azotobacter vinelandii* and *Rhodopseudomonas palustris* VnfEN cluster with NifEN and are thought to have evolved by a duplication of *nifEN* (*22*)]. The proteins comprising the VnfEN lineage depicted in Fig. 1C are encoded by genes that are proximal to VnfDK encoding genes in these four genomes.

NifB

The branching order of NifB deviates at several nodes when compared to the branching order of NifDKEN proteins. Like NifDKEN sequences, two subclusters of NifB emerge, although in contrast to the NifDKEN phylogram, these are weakly supported (posterior probabilites of 54 and 77). One subcluster contains sequences from from *Heliobacterium* spp. (*Firmicutes*) and *Desulfitobacterium hafniense* (*Firmicutes*), organisms which branch at the base of bacterial Nif cluster B2. In addition, this cluster contains NifB from *Dehalococcoides etheneogenes* (*Chloroflexi*) which together form a sister group to sequences from *Methanosarcina* spp. (*Methanaosarcinales*). NifDKEN from *D. etheneogenes* clusters in bacterial cluster B1 and NifDKEN from the *Methanosarcinales* branch at the base of this cluster Nif B1. The second primary subcluster contains early branching sequences from *Chlorobium limicola* (*Chlorobi*), *Syntrophobacter fumaroxidans* (*Proteobacteria*), and *Desulfovibrio vulgaris* (*Proteobacteria*). These taxa contain NifDKEN that branch in bacterial Nif sequence cluster B1. Also included in this subcluster are NifB from cyanobacteria and proteobacteria whose NifDKEN cluster in bacterial Nif cluster B2.

Several hypotheses are put forth to reconcile the differences in topology of the NifB lineges when compared to the individual NifDKEN lineages. The first hypothesis invokes an evolutionary trajectory for NifB that is independent of NifDKEN, which requires several early lateral gene transfer events in order to arrive at the extant NifB topology. In this hypothesis, NifB lacking the NifX domain is transferred laterally from a methanogen with character transitional between the *Methanosarcinales* and the

Methanobacteriales/*Methanococcales* to an anaerobic member of the *Firmicutes* most akin to *Heliobacterium* spp./*Desulfitobacterium hafniense* (*Firmicutes*) or to *D. etheneogenes* (*Chloroflexi*). Then, depending on the recipient of *nifB* in this lineage, a second lateral gene transfer event is required to move the sequence into the other phyla. This hypothesis would be consistent with the presence of NifB lacking NifX in all members of this sequence cluster. Independently, NifB must also have been transferred laterally from a methanogen with character transitional between the *Methanosarcinales* and the *Methanobacteriales*/*Methanococcales* to an anaerobe akin to *C. limicola*, *S. fumaroxidans*, or *D. vulgaris*. At some point following this LGT event, NifB and NifX were fused, which would explain the presence of a fused NifB-NifX in all of the derived sequences. However, this scenario is complicated by the fact that firmicutes contain both fused (*Alkaliphilus metalliredigenes* clade) and unfused (*Heliobacterium* spp. clade) NifB.

A second hypothesis derives from phylogenetic reconstructions of NifB where the first 26 positions of the N-terminus and last 40 positions of the C-terminus of the NifB alignment block are deleted (results in 283 character versus 348 character block) (data not shown). In these analyses, all of the early branching lineages containing only the SAM domain form an early branching polytomous node, with a fully resolved monophyletic lineage comprised of bacterial Nif cluster B2 sequences (excluding the *Heliobacterium* spp./*D. hafniense* lineage) emmanating from this node. Importantly, NifB from *D. etheneogenes* forms a weakly supported cluster with the *Heliobacterium* spp./*D. hafniense* lineage, which is consistent with the branching order of the individual NifDKEN lineages. Collectively, these observations are consistent with the argument

that the core sequence in early evolving lineages of NifB is highly conserved, perhaps due to the age of these lineages. The characters that are contributing the most to the branching order among the early evolving NifB lineages (all of which lack the fused NifX domain) occur in the N- and C-terminal regions of the protein, which when deleted make the phylogenetic reconstruction of the protein difficult to resolve (see above). A comparison of NifB from the two closely related strains *Methanosarcina barkerii* and *Methanosarcina acetivorans* indicates that the core NifB (283 characters) from these two strains are 100% identical whereas the full length (348 character) NifB is only 86% similar. This indicates that substitution in these regions is random. Thus, while small differences in substitutions in these regions may contribute to the deviations in the branching order of NifB from NifDKEN, the fact that proteins from methanogens branch early is consistent with these sequences universally only containing the SAM domain, suggesting this to be the ancestral state and adding confidence to the hypothesis that methanogens have retained the character most representative of the ancestral state. Importantly, the observation that NifB from the *Firmicutes* and *Proteobacteria* contain both the fused NifB-NifX and unfused, SAM-only forms of NifB, indicates that the *nifB-nifX* fusion event occurred post lateral gene transfer from the methanogens, most likely in the *Firmicutes* or *Proteobacteria*. By extension this observation would suggest that these phyla emerged later than the methanogens, a hypothesis which is consistent with the branching order of the NifDKEN and full length NifB, and also with organization of the nif operon in these taxa.

Supplemental Table A1. Accession numbers of representative sequences used in the present study. Representative sequences were selected to sample the primary lineages of each homolog, using approaches as outlined in Boyd et al., 2011 (*1*). A full list of nitrogenase and Bch homologs can be found in Boyd et al., 2011 (*1*).

Anf/Vnf/Nif/Uncharacterized Nitrogenase HDK Homologs

Taxon	NifH Homologs	NifD Homologs	NifK Homologs
Acidithiobacillus ferrooxidans ATCC 23270	YP_002219685	YP_002219684	YP_002219683
Alkaliphilus metalliredigens QYMF	YP_001321310	YP_001321307	YP_001321306
Anabaena variabilis ATCC 29413 vnf	YP_324416	YP_324526	YP_324527
Anabaena variabilis ATCC 29413 nif	YP_324741	YP_324742	YP_324743
Azotobacter vinelandii AvOP nif	YP_002797378	YP_002797379	YP_002797380
Azotobacter vinelandii AvOP anf	YP_002801975	YP_002801974	YP_002801972
Azotobacter vinelandii AvOP vnf	YP_002797502	YP_002797497	YP_002797495
Beijerinckia indica subsp. indica ATCC 9039	YP_001831615	YP_001831616	YP_001831617
Caldicellulosiruptor saccharolyticus DSM 8903	YP_001181234	YP_001181231	YP_001181230
Candidatus Azobacteroides pseudotrichonymphae CFP2	YP_002309219	YP_002309222	YP_002309223
Candidatus Desulforudis audaxviator MP104C	YP_001716343	YP_001716346	YP_001716347
Candidatus Methanosphaerula palustris E1-9c	YP_002465657	YP_002465654	YP_002465653
Chlorobium tepidum TLS	NP_662417	NP_662420	NP_662421
Chloroherpeton thalassium ATCC 35110	YP_001996732	YP_001996735	YP_001996737
Chloroherpeton thalassium ATCC 35110	YP_001995946	YP_001995943	YP_001995942
Clostridium acetobutylicum ATCC 824	NP_346894	NP_346897	NP_346898
Clostridium beijerinckii NCIMB 8052	ABR34169	ABR34172	ABR34173
Clostridium kluyveri NBRC 12016	YP_001395138	YP_001395137	YP_001395135
Dehalococcoides ethenogenes 195	YP_181872	YP_181869	YP_181868
Desulfitobacterium hafniense Y51	YP_520504	YP_520503	YP_520502
Desulfotomaculum reducens MI-1	YP_001114150	YP_001114147	YP_001114146
Desulfovibrio vulgaris subsp. vulgaris DP4	YP_009055	YP_961292	YP_009051
Geobacter lovleyi SZ	YP_001950896	YP_001950895	YP_001950894
Geobacter sp. M21	YP_003021955	YP_003021954	YP_003021953
Heliobacterium modesticaldum Ice1	YP_001679706	YP_001679707	YP_001679708
Klebsiella pneumoniae 342	YP_002237565	YP_002237564	YP_002237563
Lyngbya sp. PCC 8106	ZP_01620768	ZP_01620767	ZP_01620766
Magnetospirillum magnetotacticum MS-1	YP_420937	ZP_00054386	ZP_00054385
Methanobacterium thermoautotrophicum str. Delta H	NP_276673	NP_276676	NP_276677
Methanococcus aeolicus Nankai-3	YP_001325622	YP_001325619	YP_001325618
Methanococcus maripaludis strain S2	NP_987973	NP_987976	NP_987977
Methanococcus vannielii SB	YP_001322591	YP_001322588	YP_001322587

Methanosarcina acetivorans str. C2A	NP_618766	NP_618769	NP_618770
Methanosarcina acetivorans str. C2A	NP_616152	NP_616155	NP_616157
Methanosarcina acetivorans str. C2A	NP_616144	NP_616149	NP_616147
Methylacidiphilum infernorum V4	YP_001940528	YP_001940526	YP_001940525
Methylocella silvestris BL2	YP_002363879	YP_002363878	YP_002363877
Nostoc sp. PCC 7120	NP_485497	NP_485484	NP_485483
Opitutaceae bacterium TAV2	ZP_03723745	ZP_03723746	ZP_03723747
Pelobacter carbinolicus DSM 2380	YP_357508	YP_357509	YP_357510
Rhodobacter sphaeroides ATCC 17025	YP_001167452	YP_001167453	YP_001167454
Rhodopseudomonas palustris BisA53	YP_782790	YP_782789	YP_782787
Rhodopseudomonas palustris HaA2	YP_484590	YP_484591	YP_484592
Rhodospirillum centenum SW	YP_002299844	YP_484591	YP_002299842
Rhodospirillum rubrum ATCC 11170	YP_426483	YP_426482	YP_426480
Roseiflexus castenholzii DSM 13941	YP_001434094	YP_001434092	YP_001434091
Roseiflexus sp. RS-1	YP_001275558	YP_001275556	YP_001275555
Sinorhizobium medicae WSM419	YP_001314762	YP_001314761	YP_001314760
Synechococcus sp. JA-2-3B'a	YP_475238	YP_476681	YP_476682
Syntrophobacter fumaroxidans MPOB	YP_845148	YP_845145	YP_845144
Thermodesulfovibrio yellowstonii DSM 11347	YP_002249508	YP_002249507	YP_002249506
Zymomonas mobilis subsp. mobilis ZM4	ZP_04760608	YP_163559	YP_163560
Methanocaldococcus infernus ME	YP_003615677	YP_003615674	YP_003615673
Syntrophothermus lipocalidus DSM 12680	YP_003703438	YP_003703435	YP_003703434
Methanocaldococcus vulcanius M7	YP_003247421	YP_003247424	YP_003247425
Methanosarcina mazei strain Goe1	NP_632743	NP_632746	NP_632747
Methanocaldococcus sp. FS406-22	YP_003457468	YP_003457471	YP_003457472
Methanothermococcus okinawensis IH1	ZP_07330063	ZP_07330060	ZP_07330059
Oscillochloris trichoides	ZP_07684114	ZP_07684112	ZP_07684111
Chlorothrix halophila	NA*	NA*	NA*
Anabaena variabilis ATCC 29413	YP_322845	YP_322843	YP_323968
Chlorobium limicola DSM 245	YP_001944195	YP_001944197	YP_001944196

Nfl Homologs

Taxon	NflH Homologs	NflD Homologs
Methanobacterium thermoautotrophicum	AAB85148	AAB85997
Methanobrevibacter smithii	YP_001274280	YP_001273733
Methanocaldococcus jannaschii	NP_247874	NP_248427
Methanocaldococcus sp. FS406-22	YP_003458692	YP_003458054
Methanococcoides burtonii	YP_565723	YP_565722
Methanococcus aeolicus	YP_001325412	YP_001325501
Methanococcus maripaludis C5	YP_001098039	YP_001097721
Methanococcus maripaludis C6	YP_001548852	YP_001548533
Methanococcus maripaludis C7	YP_001330364	YP_001330639
Methanococcus maripaludis S2	CAF29703	CAF29984

Methanococcus vannielii	YP_001323663	YP_001323921	
Methanocorpusculum labreanum	YP_001029608	YP_001029961	
Methanopyrus kandleri	AAM02629	AAM02598	
Methanosaeta thermophila	YP_843600	YP_842548	
Methanosarcina acetivorans	AAM05043	AAM06983	
Methanosarcina barkeri	YP_303736	YP_303910	
Methanosarcina mazei	AAM30210	AAM30211	
Methanosphaera stadtmanae	YP_448145	YP_448474	

Chl/Bch Homologs

Taxon	ChlL/BchL Homologs	ChlN/BchN Homologs	ChlB/BchB
Bradyrhizobium sp. BTAi1	YP_001242230	YP_001242233	YP_001242232
Methylobacterium populi BJ001	YP_001927985	YP_001927988	YP_001927987
Erythrobacter sp. NAP1	ZP_01041680	ZP_01041677	ZP_01041678
Hoeflea phototrophica DFL-43	ZP_02167537	ZP_02167540	ZP_02167539
Halorhodospira halophila SL1	YP_001003200	YP_001003203	YP_001003202
Prochlorococcus sp. CC9311	YP_731178	YP_731176	YP_731177
Synechococcus sp. RCC307	YP_001227822	YP_001227820	YP_001227821
Prochlorococcus marinus MIT 9515	YP_001010923	YP_001010925	YP_001010924
Roseobacter sp. AzwK-3b	ZP_01902759	ZP_01902756	YP_002464757
Chloroflexus aggregans DSM 9485	YP_002464754	YP_002464757	YP_002464756
Roseiflexus castenholzii HLO8, DSM 13941	YP_001431649	YP_001431647	YP_001431648
Chlorobium phaeobacteroides DSM 266	YP_912798	YP_912800	YP_912799
Cyanothece sp. CCY 0110	ZP_01730952	ZP_01730955	ZP_01728917
Synechococcus sp. JA-2-3B	YP_477257	YP_477255	YP_478401
Gloeobacter violaceus PCC 7421	NP_925316	NP_925315	NP_923161
Heliobacterium modesticaldum Ice1	YP_001679876	YP_001679877	YP_001679878
Anabaena variabilis ATCC 29413	YP_322845	YP_322843	YP_323968
Chlorobium limicola DSM 245	YP_001944195	YP_001944197	YP_001944196

Supplemental Table A2. Accession numbers for nitrogenase deduced amino acid sequences used in this study.

Taxon (gene description)[a]	NifD	NifK	NifE	NifN	NifB
Acidithiobacillus ferrooxidans ATCC 23270	NC_011761	YP_002425948	YP_002425945	YP_002425944	YP_002425957
Alkaliphilus metalliredigenes QYMF	YP_001321307	YP_001321306	YP_001321305	YP_001321304	YP_001321303
Anabaena variabilis 29413 *nif*	YP_324742	YP_324743	YP_324745	YP_324745	YP_324738
Anabaena variabilis 29413 *anf*	YP_324526	YP_324527	YP_324528	YP_324529	
Anaeromyxobacter Fw109-5	YP_001380210	YP_001380209	YP_001380208	YP_001380208	YP_001380205
Azoarcus BH72	YP_932043	YP_932044	YP_932066	YP_932065	YP_932027
Azorhizobium caulinodans ORS 571	YP_001523956	YP_001523955	YP_001523954	YP_001523953	YP_001526330
Azotobacter vinelandii AvOP *nif*	ZP_00417578	ZP_00417579	ZP_00417584	ZP_00417585	ZP_00419328
Azotobacter vinelandii AvOP v*nf*	ZP_00416683	ZP_00416685	ZP_00416673	ZP_00417584	
Azotobacter vinelandii AvOP *anf*	ZP_00417194	ZP_00417192			
Bradyrhizobium BTAi1	YP_001241771	YP_001241770	YP_001241769	YP_001241768	YP_001241752
Bradyrhizobium japonicum 110	NP_768383	NP_768384	NP_768385	NP_768386	NP_768399
Bradyrhizobium ORS278	YP_001207333	YP_001207332	YP_001207331	YP_001207330	YP_001207314
Burkholderia phymatum STM 815	YP_001863690	YP_001863691	YP_001863665	YP_001863666	YP_001863678
Burkholderia vietnamiensis G4	YP_001115196	YP_001115197	YP_001115200	YP_001115201	YP_001115174
Burkholderia xenovorans LB400	YP_553848	YP_553847	YP_553844	YP_553843	YP_553868
Caldicellulosiruptor saccharolyticus DSM 8903	YP_001181231	YP_001181230	ABP68038		YP_001181228
Candidatus Methanoregula boonei 6A8	YP_001404306	YP_001404307	YP_001404308	YP_001404309	YP_001404309
Chlorobium chlorochromatii CaD3	YP_379551	YP_379552	YP_379553	YP_379554	YP_379555
Chlorobium ferrooxidans DSM 13031	ZP_01385046	ZP_01385047	ZP_01385048	ZP_01385049	ZP_01385050
Chlorobium limicola DSM245	YP_001942744	YP_001942743	YP_001942742	YP_001942741	YP_001942740
Chlorobium phaeobacteroides BS1	YP_001960153	YP_001960154	YP_001960155	YP_001960156	YP_001960157
Chlorobium phaeobacteroides DSM 266	YP_911218	YP_911217	ABL64793	YP_911216	YP_911215
Chlorobium tepidum TLS	NP_662420	NP_662421	NP_662422	NP_662423	NP_662424
Clostridium acetobutylicum ATCC 824	NP_346897	NP_346898	NP_346899	NP_346900	NP_346900
Clostridium beijerincki NCIMB 8052	YP_001309128	YP_001309129	YP_001309130	YP_001309131	YP_001309131

Clostridium botulinum A 19397	YP_001383075	YP_001383076	YP_001383075	YP_001383076	YP_001384959
Clostridium kluyveri DSM 555 *nif*	YP_001396457	YP_001396456	YP_001396455	YP_001396454	YP_001396454
Clostridium kluyveri DSM 555 *vnf*	YP_001395137	YP_001395135	YP_001395134	YP_001395133	
Clostridium kluyveri DSM 555 *anf*	YP_001393772	YP_001393774			
Clostridium thermocellum ATCC 27405	YP_001037984	YP_001037985			YP_001037987
Crocosphaera watsonii WH 8501	ZP_00516387	ZP_00516388	ZP_00516390	ZP_00516391	ZP_00516383
Cyanothece CCY0110	ZP_01727766	ZP_01727767	ZP_01727769	ZP_01727770	ZP_01727760
Dechloromonas aromatica DCB-2	YP_284633	YP_284632	YP_284720	YP_284724	YP_284673
Dehalococcoides ethenogenes 195	YP_181869	YP_181868	YP_181867	YP_181866	YP_181862
Desulfitobacterium hafniense DCB-2	ZP_01372575	ZP_01372574	ZP_01372573	ZP_01372572	ZP_01372569
Desulfitobacterium hafniense Y51	YP_520503	YP_520502	YP_520501	YP_520500	YP_520498
Desulfotomaculum reducens MI-1	YP_001114147	YP_001114146	YP_001114145	YP_001114144	YP_001114143
Desulfovibrio vulgaris DP4	YP_961292	YP_961293	YP_961295	YP_961296	YP_961297
Desulfovibrio vulgaris Hildenborough	YP_009052	YP_009051	YP_009049	YP_009048	YP_009047
Desulfuromonas acetoxidans DSM 684	ZP_01312342	ZP_01312341	ZP_01313365	ZP_01313366	
Erwinia carotovora atroseptica SCRI1043	YP_051046	YP_051045	YP_051040	YP_051039	YP_051028
Frankia alni ACN14A	YP_716938	YP_716937	YP_716936	YP_716935	YP_716929
Frankia CcI3	YP_483562	YP_483561	YP_483560	YP_483559	YP_483553
Frankia EAN1pec	YP_001511120	YP_001511119	YP_001511118	YP_001511117	YP_001511111
Geobacter bemidjiensis BEM	YP_002138884	YP_002138885	YP_002138886	YP_002138886	YP_002138892
Geobacter lovleyi SZ	YP_001950895	YP_001950894	YP_001950885	ZP_01591999	
Geobacter metallireducens GS-15	YP_383630	YP_383631	YP_383636	YP_383636	ABB30923
Geobacter sulfurreducens PCA	NP_953864	NP_953863	NP_953860	NP_953860	AAR36193
Geobacter uraniumreducens Rf4	YP_001229953	YP_001229954	YP_001229978	YP_001229978	ABQ25413
Halorhodospira halophila SL1	YP_001001869	YP_001001868	YP_001001859	YP_001001860	YP_001001877
Heliobacterium chlorum	BAD95754	BAD95755	BAD95756	BAD95757	BAD95760
Heliobacterium modesticaldum Ice1	YP_001679707	YP_001679708	YP_001679709	YP_001679710	YP_001679713
Klebsiella pneumoniae 78578	CAA31667	CAA31668	CAA31671	CAA31672	CAA31683

Magnetococcus MC-1	YP_865118	YP_865117	YP_865112	YP_865111	YP_865123
Magnetospirillum magneticum AMB-1	YP_420936	YP_420935	YP_420932	YP_420931	BAE50387
Mesorhizobium loti MAFF303099	NP_106490	NP_106491	NP_106492	NP_106493	NP_106447
Methanococcus aeolicus Nankai-3	YP_001325619	YP_001325618	YP_001325615	YP_001325614	ABR56171
Methanococcus maripaludis C5	YP_001097187	YP_001097188	YP_001097189	YP_001097190	ABO35223
Methanococcus maripaludis C6	YP_001549843	YP_001549842	YP_001549841	YP_001549840	ABX01050
Methanococcus maripaludis C7	YP_001329322	YP_001329323	YP_001329324	YP_001329325	ABR66740
Methanococcus maripaludis S2	CAF30412	CAF30413	CAF30414	CAF30415	CAF30214
Methanococcus vannielii SB	YP_001322588	YP_001322587	YP_001322586	YP_001322585	ABR55426
Methanosarcina acetivorans C2A *nif*	AAM07249	AAM07250	AAM07251	AAM07252	AAM07541
Methanosarcina acetivorans C2A *vnf*	AAM04635	AAM04637	AAM04638	AAM04639	
Methanosarcina acetivorans C2A *anf*	NP_616149.1	NP_616147.1			
Methanosarcina barkeri str. Fusaro *nif*	YP_303733	YP_303732	YP_303731	YP_303730	AAZ69118
Methanosarcina barkeri str. Fusaro *vnf*	YP_305783	YP_305781	YP_305779	YP_305778	
Methanosarcina barkeri str. Fusaro *anf*	YP_305080.1	YP_305078.1			
Methanosarcina mazei Go1	AAM30418	AAM30419	AAM30420	AAM30421	AAM30453 / AAM30455
Methanothermobacterium thermautotrophicum Delta H	AAB86037	AAB86038	AAB86039	AAB86040	AAB86337
Methylococcus capsulatus str. Bath	YP_112765	YP_112766	YP_112768	YP_112769	YP_112740
Nodularia spumigena CCY9414	ZP_01628430	ZP_01628425	ZP_01628423	ZP_01628422	
Nostoc PCC 7120	NP_485484	NP_485483	NP_485481	NP_485480	BAB77883
Nostoc punctiforme PCC 73102	ZP_01111244	ZP_01111245	ZP_01112340	ZP_01112341	
Pelobacter carbinolicus DSM 2380	YP_357509	YP_357510	YP_355743	YP_355744	YP_357517
Pelobacter propionicus DSM 2379	YP_903118	YP_903119	YP_903141 (E/N fusion)	YP_903141	ABL01056
Pelodictyon luteolum DSM 273	YP_375431	YP_375432	YP_375433	YP_375434	YP_375435
Pelodictyon phaeoclathratiforme BU-1	ZP_00588540	ZP_00588541	ZP_00588542	ZP_00588543	ZP_00588544
Polaromonas naphthalenivorans CJ2	YP_982572	YP_982571	YP_982554	YP_982553	YP_981809
Prosthecochloris aestuarii DSM 271	ZP_0059	ZP_0059	ZP_0059	ZP_0059	ZP_0059

	2298	2299	2300	2301	2302
Prosthecochloris vibrioformis DSM 265	YP_001130860	YP_001130861	YP_001130862	YP_001130863	YP_001130864
Pseudomonas stutzeri A1501	YP_001171864	YP_001171865	YP_001171870	YP_001171871	
Rhizobium etli CFN42	NP_659837	NP_659838	NP_659839	NP_659840	NP_659809
Rhizobium leguminosarum vicae 3841	YP_770439	YP_770438	YP_770437	YP_770436	CAD58674
Rhodobacter sphaeroides 17025	YP_001167453	YP_001167454	YP_001167455	YP_001167456	YP_001167447
Rhodobacter sphaeroides 17029	YP_001044066	YP_001044065	YP_001044064 (E/N fusion)	YP_001044064	YP_001044072
Rhodobacter sphaeroides 241	YP_353613	YP_353612	YP_353611	YP_353610	YP_353619
Rhodopseudomonas palustris A53	YP_783432	YP_783431	YP_783430	YP_783429	YP_783441
Rhodopseudomonas palustris B18	YP_534304	YP_534303	YP_534302	YP_534301	YP_534314
Rhodopseudomonas palustris B5	YP_568213	YP_568214	YP_568215	YP_568216	YP_568202
Rhodopseudomonas palustris CGA009 *nif*	NP_949953	NP_949952	NP_949951	NP_949950	NP_949964
Rhodopseudomonas palustris CGA009 *vnf*	NP_946728	NP_946730	NP_946723	NP_946722	
Rhodopseudomonas palustris CGA009 *anf*	YP_001990626	YP_001990624			
Rhodopseudomonas palustris HaA2	YP_484591	YP_484592	YP_484593	YP_484594	YP_484580
Rhodospirillum rubrum ATCC 11170 *nif*	YP_426099	YP_426100	YP_427373	YP_427372	YP_426082
Rhodospirillum rubrum ATCC 11170 *anf*	YP_426482	YP_426480			
Roseiflexus castenholzii DSM 13941	YP_001434092	YP_001434091			YP_001434093
Roseiflexus RS-1	YP_001275556	YP_001275555			YP_001275557
Sinorhizobium medicae WSM419	YP_001314761	YP_001314760	YP_001314759	YP_001314731	YP_001314769
Sinorhizobium meliloti 1021	NP_435696	NP_435697	NP_435698	NP_435724	NP_435688
Synechococccus JA-3-3B'a(2-13)	YP_476681	YP_476682	YP_476670	YP_476669	YP_476676
Synechococcus sp. JA-3-3Ab	YP_475237	YP_475236	YP_475248	YP_475249	YP_475242
Syntrophobacter fumaroxidans MPOB	YP_845145	YP_845144	YP_845142	YP_845141	YP_845140
Syntrophomonas wolfei Goettingen	YP_753116	YP_753115			
Trichodesmium erythraeum IMS101	YP_723618	YP_723619	YP_723620 (E/N fusion)	YP_723620	YP_723614
Wolinella succinogenes DSM 1740	NP_907558	NP_907557	NP_907556	NP_907555	NP_907563

Xanthobacter autotrophicus Py2	YP_0014 15005	YP_0014 15006	YP_0014 15007	YP_0014 15008	YP_0014 15027
Zymomonas mobilis ZM4	YP_1635 59.1	YP_1635 60.1	YP_1635 61.1	YP_1635 62.1	AAV904 41

[a]Alternative nitrogenase are indicated in italics following the description of the taxon

Supp. Table A3. Accession numbers for BchNBYZ loci used in this study.

Taxon	BchN	BchB	BchY	BchZ
Acaryochloris marina MBIC11017	YP_001515786.1	YP_001515878.1		
Anabaena variabilis ATCC 29413	YP_322843.1	YP_323968.1		
Bradyrhizobium sp. BTAi1	YP_001242233.1	YP_001242232.1	YP_001242252.1	YP_001242251.1
Bradyrhizobium sp. ORS278	YP_001203754.1	YP_001203755.1	YP_001203735.1	YP_001203736.1
Chlorobaculum parvum NCIB 8327	YP_001997846.1	YP_001997847.1	YP_001997999.1	YP_001997747.1
Chlorobium chlorochromatii CaD3	YP_380109.1	YP_380108.1	YP_378640.1	YP_380010.1
Chlorobium ferrooxidans DSM 13031	ZP_01386400.1	ZP_01386401.1	ZP_01385227.1	ZP_01386940.1
Chlorobium limicola DSM 245	YP_001944197.1	YP_001944196.1	YP_001944043.1	YP_001942235.1
Chlorobium phaeobacteroides DSM 266	YP_001960645.1	YP_912799.1	YP_001960453.1	YP_001958629.1
Chlorobium phaeobacteroides BS1	YP_912800.1	YP_001960644.1	YP_912511.1	YP_912912.1
Chlorobium tepidum TLS	NP_663026.1	NP_663025.1	NP_662705.1	NP_662999.1
Chloroflexus aggregans DSM 9485	YP_002464757.1	YP_002464756.1	YP_002462207.1	YP_002462208.1
Chloroflexus aurantiacus J-10-fl	YP_001636152.1	YP_001636151.1	YP_001637372.1	YP_001637373.1
Chloroflexus sp. Y-400-fl	Q9F6X6.1	YP_002570475.1	YP_001637372.1	YP_001637373.1
Chloroherpeton thalassium ATCC 35110	YP_001996283.1	YP_001996282.1	YP_001995116.1	YP_001995732.1
Crocosphaera watsonii WH 8501	ZP_00515910.1	ZP_00518641.1		
Cyanothece sp. ATCC 51142	YP_001805948.1	YP_001803370.1		
Cyanothece sp. CCY0110	ZP_01730955.1	ZP_01728917.1		
Cyanothece sp. PCC 7424	YP_002377662.1	YP_002375695.1		
Cyanothece sp. PCC 7425	YP_002484618.1	YP_002484347.1		
Cyanothece sp. PCC 7822	ZP_03155961.1	ZP_03155438.1		
Cyanothece sp. PCC 8801	YP_002371027.1	YP_002373046.1		
Cyanothece sp. PCC 8802	ZP_03142279.1	ZP_03145324.1		
Dinoroseobacter shibae DFL 12	YP_001534868.1	YP_001534869.1	YP_001534852.1	YP_001534853.1
Erythrobacter sp. NAP1	ZP_01041677.1	ZP_01041678.1	ZP_01041662.1	ZP_01041663.1
Fulvimarina pelagi HTCC2506	ZP_01440452.1	ZP_01440451.1	ZP_01440432.1	ZP_01440433.1
Gloeobacter violaceus PCC 7421	NP_925315	NP_923161		

	.1	.1		
Halorhodospira halophila SL1	YP_001003203.1	YP_001003202.1	YP_001003175.1	YP_001003174.1
Heliobacterium modesticaldum Ice1	YP_001679877.1	YP_001679878.1	YP_001679909.1	YP_001679908.1
Hoeflea phototrophica DFL-43	ZP_02167540.1	ZP_02167539.1	ZP_02167518.1	ZP_02167519.1
Jannaschia sp. CCS1	YP_508103.1	YP_508102.1	YP_508120.1	YP_508119.1
Loktanella vestfoldensis SKA53	ZP_01002311.1	ZP_01002312.1	ZP_01002374.1	ZP_01002375.1
Lyngbya sp. PCC 8106	ZP_01624540.1	ZP_01620453.1		
Methylobacterium chloromethanicum CM4	YP_002423973.1	YP_002423972.1	YP_353336.1	YP_002421726.1
Methylobacterium extorquens PA1	YP_001642255.1	YP_001642254.1	YP_001640193.1	YP_001640194.1
Methylobacterium populi BJ001	YP_001927988.1	YP_001927987.1	YP_001925543.1	YP_001925544.1
Methylobacterium radiotolerans JCM 2831	YP_001754526.1	YP_001754525.1	YP_001755484.1	YP_001755483.1
Methylobacterium sp. 4-46	YP_001770532.1	YP_001770533.1	YP_001770513.1	YP_001770514.1
Methylocella silvestris BL2	YP_002362344.1	YP_002362343.1		
Microcystis aeruginosa NIES-843	YP_001656639.1	YP_001660400.1		
Nodularia spumigena CCY9414	ZP_01632226.1	ZP_01628967.1		
Nostoc sp. PCC 7120	YP_001867733.1	YP_001868220.1		
Nostoc punctiforme PCC 73102	NP_489116.1	NP_487481.1		
Pelodictyon luteolum DSM 273	YP_374146.1	YP_374147.1	YP_374273.1	YP_374057.1
Pelodictyon phaeoclathratiforme BU-1	YP_002017274.1	YP_002017275.1	YP_002019216.1	YP_002017132.1
Prochlorococcus marinus str. AS9601	YP_001008994.1	YP_001008993.1		
Prochlorococcus marinus str. MIT 9211	YP_001550432.1	YP_001550431.1		
Prochlorococcus marinus str. MIT 9215	YP_001483827.1	YP_001483826.1		
Prochlorococcus marinus str. MIT 9301	YP_001090795.1	YP_001090794.1		
Prochlorococcus marinus str. MIT 9303	YP_001016811.1	YP_001016810.1		
Prochlorococcus marinus str. MIT 9312	YP_397042.1	YP_397041.1		
Prochlorococcus marinus str. MIT 9313	NP_895044.1	NP_895045.1		
Prochlorococcus marinus str. MIT 9515	YP_001010925.1	YP_001010924.1		
Prochlorococcus marinus str. NATL1A	YP_001014425.1	YP_001014426.1		

Prochlorococcus marinus str. NATL2A	YP_293064.1	YP_293065.1		
Prochlorococcus marinus subsp. marinus	NP_874939.1	NP_874938.1		
Prochlorococcus marinus subsp. pastoris	NP_892663.1	NP_892662.1		
Prosthecochloris aestuarii DSM 271	YP_002016686.1	YP_002016685.1	YP_002016522.1	YP_002014874.1
Prosthecochloris vibrioformis DSM 265	YP_001129805.1	YP_001129806.1	YP_001129931.1	YP_001129716.1
Rhodobacter sphaeroides 2.4.1	YP_353359.1	YP_353360.1	YP_353336.1	YP_353335.1
Rhodobacter sphaeroides ATCC 17025	YP_001167216.1	YP_001167215.1	YP_001168231.1	YP_001168230.1
Rhodobacter sphaeroides ATCC 17029	YP_001043804.1	YP_001043805.1	YP_001043781.1	YP_001043780.1
Rhodopseudomonas palustris BisA53	YP_780279.1	YP_780280.1	YP_780260.1	YP_780261.1
Rhodopseudomonas palustris BisB18	YP_531194.1	YP_531195.1	YP_531175.1	YP_531176.1
Rhodopseudomonas palustris BisB5	YP_570858.1	YP_570857.1	YP_570877.1	YP_570876.1
Rhodopseudomonas palustris CGA009	NP_946888.1	NP_946889.1	NP_946871.1	NP_946872.1
Rhodopseudomonas palustris HaA2	YP_487585.1	YP_487584.1	YP_487604.1	YP_487603.1
Rhodopseudomonas palustris TIE-1	YP_001990733.1	YP_001990734.1	YP_001990714.1	NP_946872.1
Rhodospirillum rubrum ATCC 11170	YP_425714.1	YP_425713.1	YP_428061.1	YP_428060.1
Roseiflexus castenholzii DSM 13941	YP_001431647.1	YP_001431648.1	YP_001433802.1	YP_001433801.1
Roseiflexus sp. RS-1	YP_001276249.1	YP_001276248.1	YP_001277568.1	YP_001277567.1
Roseobacter denitrificans OCh 114	YP_680560.1	YP_680559.1	YP_680531.1	YP_680530.1
Roseobacter litoralis Och 149	ZP_02142857.1	ZP_02142858.1	ZP_02142890.1	ZP_02142891.1
Roseobacter sp. AzwK-3b	ZP_01902756.1	ZP_01902757.1	ZP_01902778.1	ZP_01902777.1
Roseobacter sp. CCS2	ZP_01750124.1	ZP_01750123.1	ZP_01749903.1	ZP_01749902.1
Roseovarius sp. 217	ZP_01034606.1	ZP_01034607.1	ZP_01034628.1	ZP_01034627.1
Roseovarius sp. TM1035	ZP_01879029.1	ZP_01879030.1	ZP_01879051.1	ZP_01879050.1
Synechococcus elongatus PCC 6301	YP_170846.1	YP_172966.1		
Synechococcus elongatus PCC 7942	YP_400437.1	YP_400855.1		
Synechococcus sp. RS9916	YP_377621.1	ZP_01472612.1		
Synechococcus sp. RS9917	YP_731176.1	ZP_01079787.1		
Synechococcus sp. BL107	YP_381071.1	ZP_01469483.1		

Synechococcus sp. CC9311	YP_377621.1	YP_731177.1
Synechococcus sp. CC9605	YP_381071.1	YP_381070.1
Synechococcus sp. CC9902	YP_377621.1	YP_377622.1
Synechococcus sp. PCC 7002	YP_001735578.1	YP_001734898.1
Synechococcus sp. RCC307	YP_001227820.1	YP_001227821.1
Synechococcus sp. JA-2-3B'a(2-13)	YP_477255.1	YP_478401.1
Synechococcus sp. JA-3-3Ab	YP_475773.1	YP_474361.1
Synechococcus sp. WH 5701	ZP_01085321.1	ZP_01085322.1
Synechococcus sp. WH 7803	YP_001224395.1	YP_001224394.1
Synechococcus sp. WH 7805	ZP_01123233.1	ZP_01123232.1
Synechococcus sp. WH 8102	NP_897814.1	NP_897815.1
Synechocystis sp. PCC 6803	NP_442934.1	NP_442044.1
Thermosynechococcus elongatus BP-1	NP_683135.1	NP_683182.1
Trichodesmium erythraeum IMS101	YP_721284.1	YP_723330.1

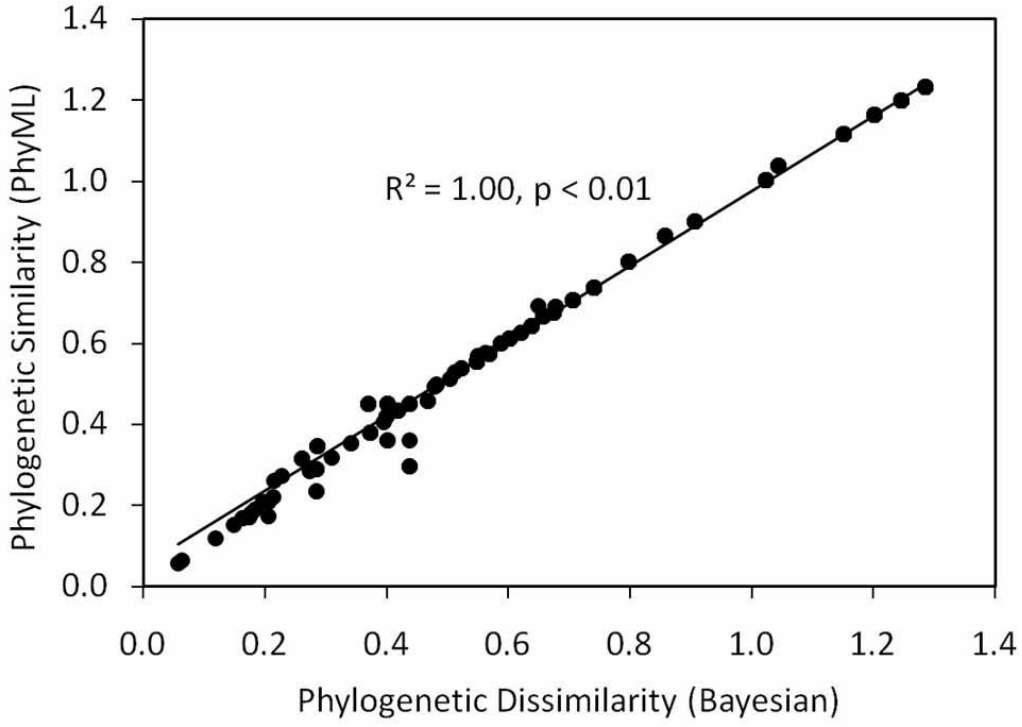


Figure A1. Plot of a Mantel regression of a matrix describing the Rao phylogenetic dissimilarity of concatenated HDK homologs inferred by PhyML as a function of the Rao phylogenetic dissimilarity of concatenated HDK homologs inferred by MrBayes. The strong positively trending correlation suggests that the topologies of the two trees are congruent.

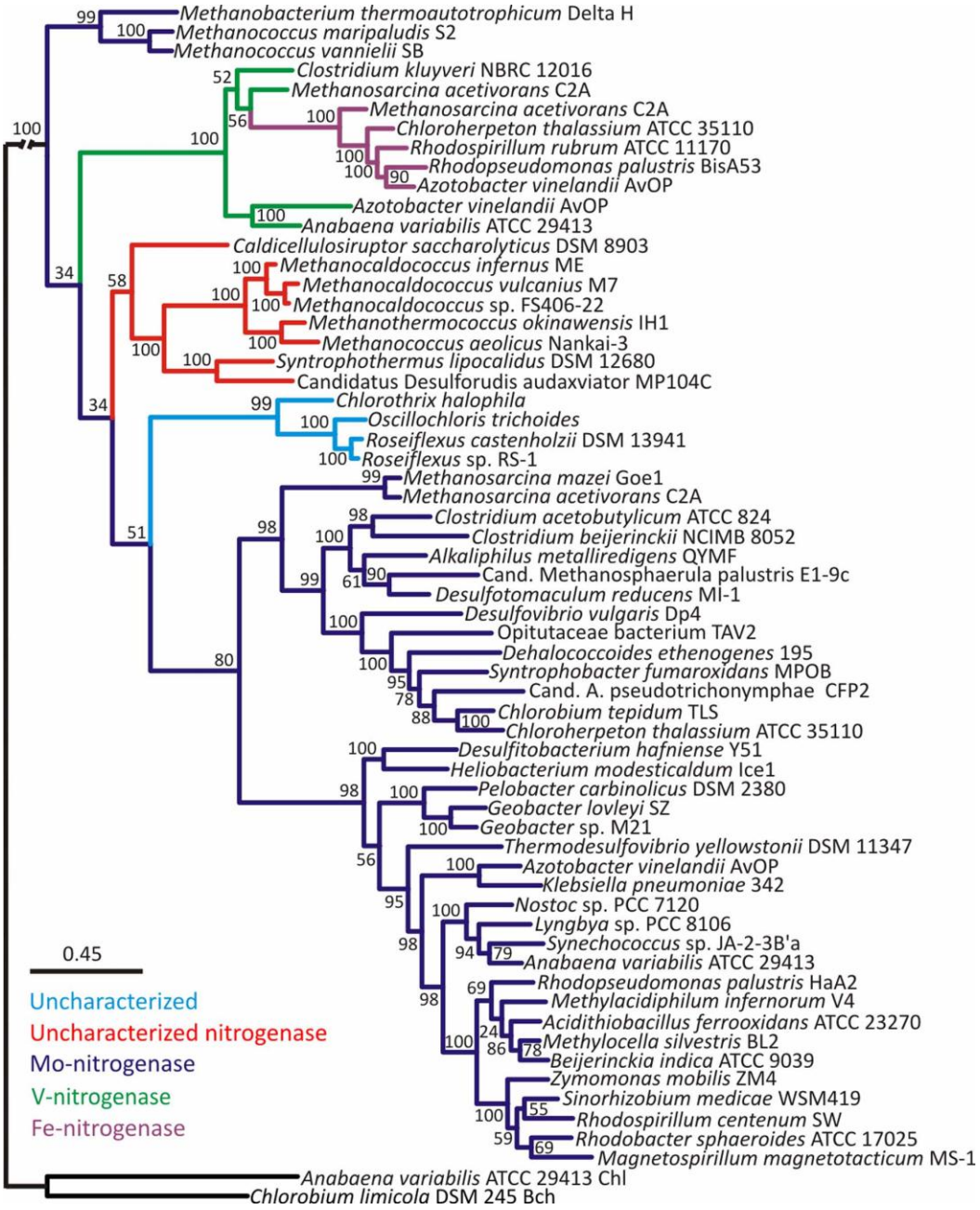


Figure A2. Maximum likelihood inferred phylogenetic tree of concatenated HDK homologs and operon structure. Bootstrap values based on 100 replicates are indicated at each node. Branches are colored dark blue (Mo-nitrogeanse, Nif), green (V-nitrogenase, Vnf), purple (Fe-nitrogenase, Anf), red (uncharacterized nitorgenase), and light blue (uncharacterized homolog). The hash at the root was introduced to conserve space.

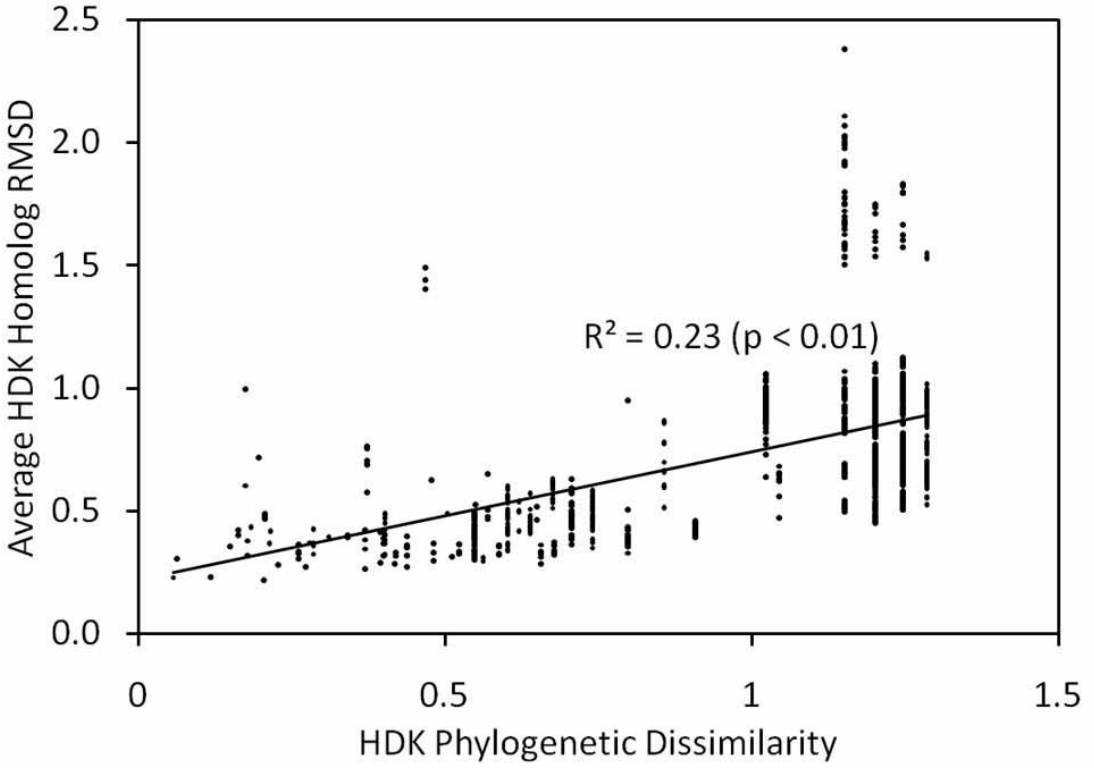


Figure A3. Plot of a Mantel regression of a matrix describing the average RMSD for H, D, and K protein structures inferred using homology modeling as a function of the Rao phylogenetic dissimilarity of concatenated HDK homologs inferred by MrBayes. The strong correlation suggests a relationship between the evolution of sequences and their inferred structures, implying that the HDK structure is evolving. The slope of the line linear regression (~ 2) suggests that the evolution of protein structure is constrained to a greater extent than the evolution of the primary sequences.

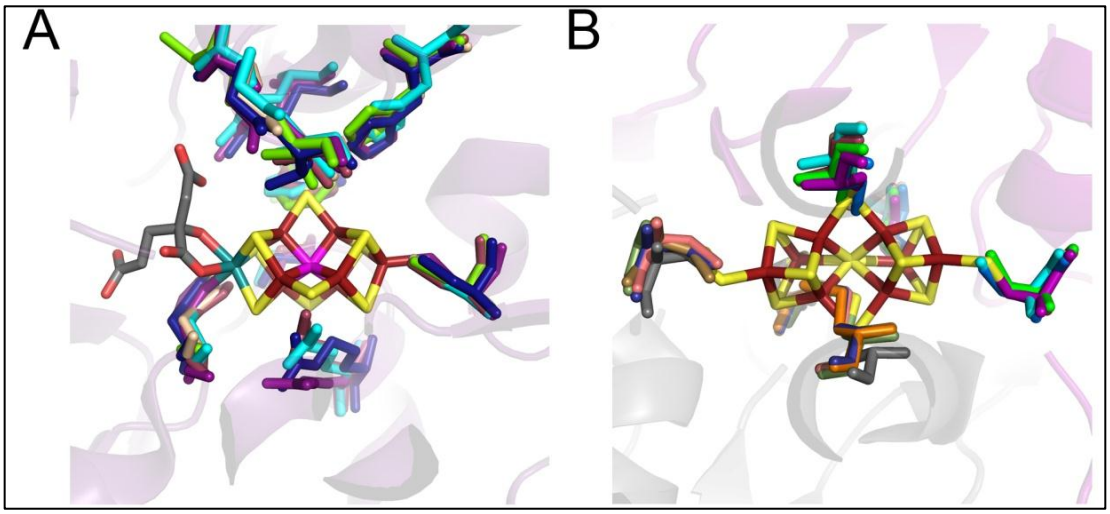


Figure A4. Structural alignment of the inferred structures of DK homologs indicating conservation in the active site (A) and P-cluster binding cavity (B). Ribbon diagram of the superimposition of NifDK from *Azotobacter vinelandii* AvOP (D, violet and K, grey), NifDK from *Methanococcus maripaludis* strain S2 (D, wheat and NifK, blue), UncDK from *Methanocaldococcus* sp. FS406-22 (D, cyan and K, orange), UncDK from *Roseiflexus* sp. RS-1 (D, marine, and K, sand), VnfDK from *Methanosarcina acetivorans* str. C2A (D, raspberry, and K, pale green), and AnfDK from *Azotobacter vinelandii* AvOP (D, green, and K, salmon), with the FeMo-co (Panel A) and P-cluster (Panel B) depicted as stick representations. Dark red, Fe; yellow, S; grey, C; red, O; teal, Mo; unknown, magenta. Protein Data Bank ID for *Azotobacter vinelandii* AvOP 1MIN.

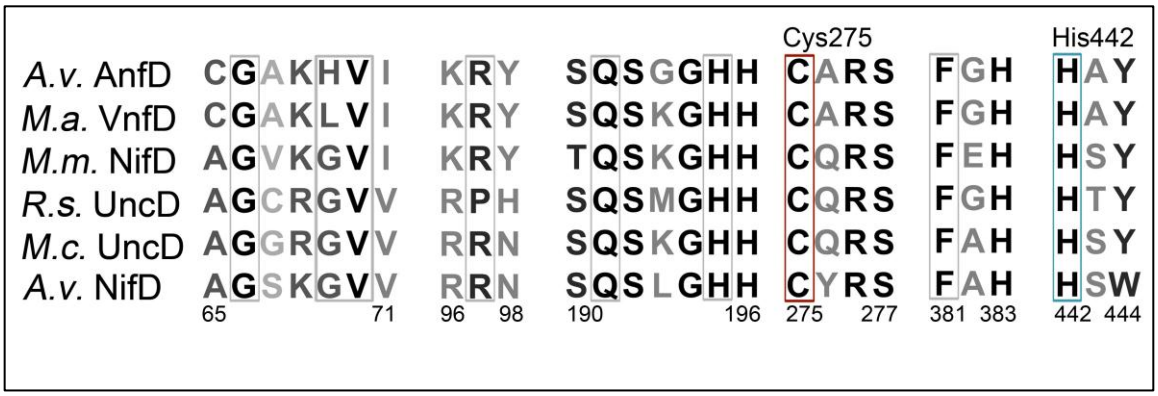


Figure A5. Amino acid sequence conservation in selected residues that ligate FeMo-co (Cys275, red box, and His442, blue box) and that have been implicated as important in the FeMo-co binding pocket (indicated by a gray box). Representative NifD, VnfD, AnfD, and UncD (uncharacterized nitrogenase). Numbering is based on NifD from *Azotobacter vinelandii* AvOP. Abbreviations: A.v., *Azotobacter vinelandii* AvOP; M.a., *Methanosarcina acetivorans* str. C2A; M.m., *Methanococcus maripaludis* strain S2; R.s., *Roseiflexus* sp. RS-1; M.c. *Methanocaldococcus* sp. FS406-22. The conservation in the active site environment and active site cluster ligands between classes of nitrogenase suggest that once the active site cavity evolved, it was maintained through time.

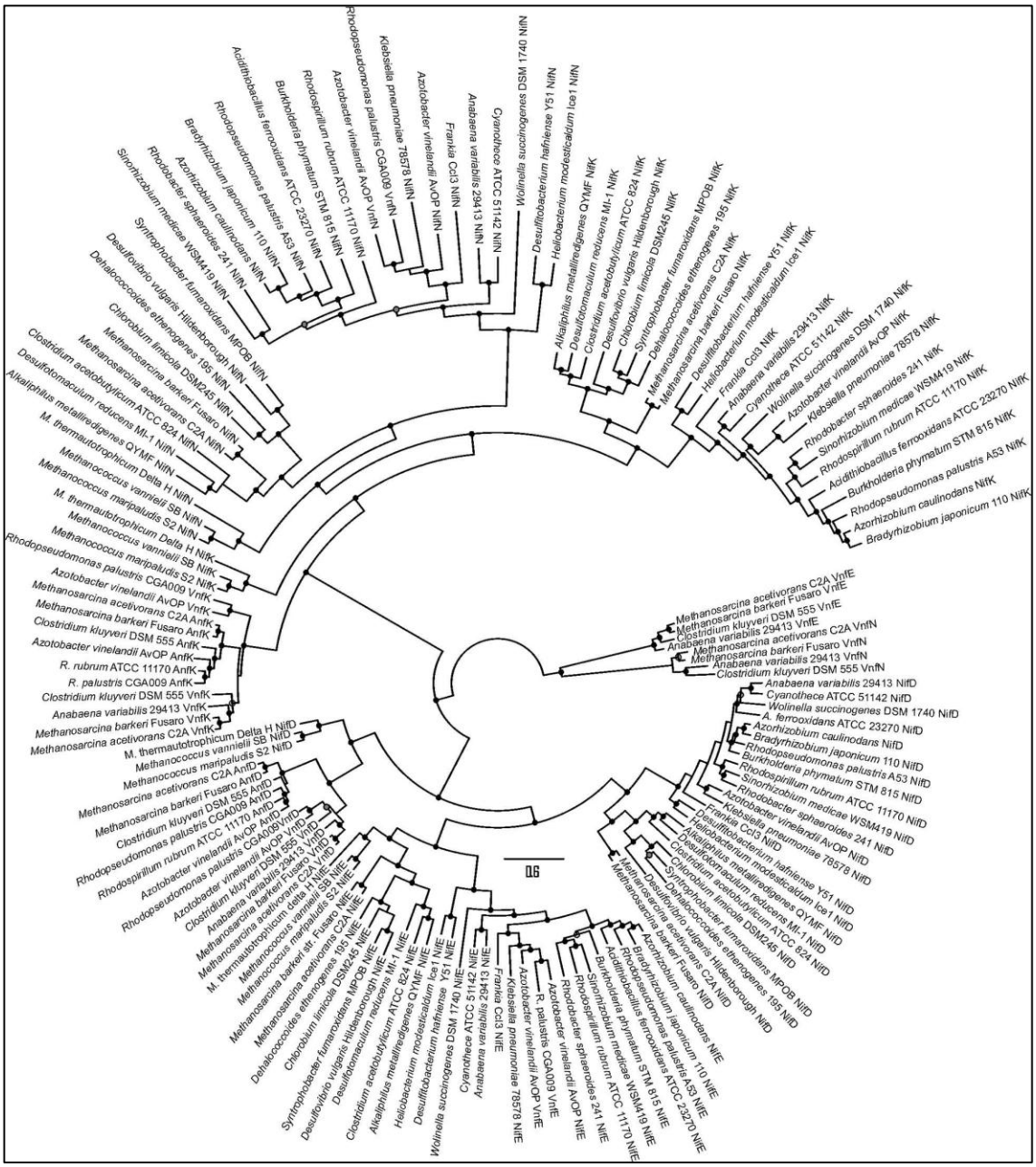


Figure A6. Bayesian consensus phylogenetic tree of Nif/Vnf/AnfD, K, E, and N proteins from 40 taxa that represent the primary lineages for each of the 4 loci. Black circles at nodes denote >90% posterior probability (PP), gray circles at nodes denote >80% PP, open circles denote >70% PP, and no symbol denotes 50 to 70% PP. Nodes with <50% PP were collapsed; barequals 6 substitutions per 10 sequence positions.

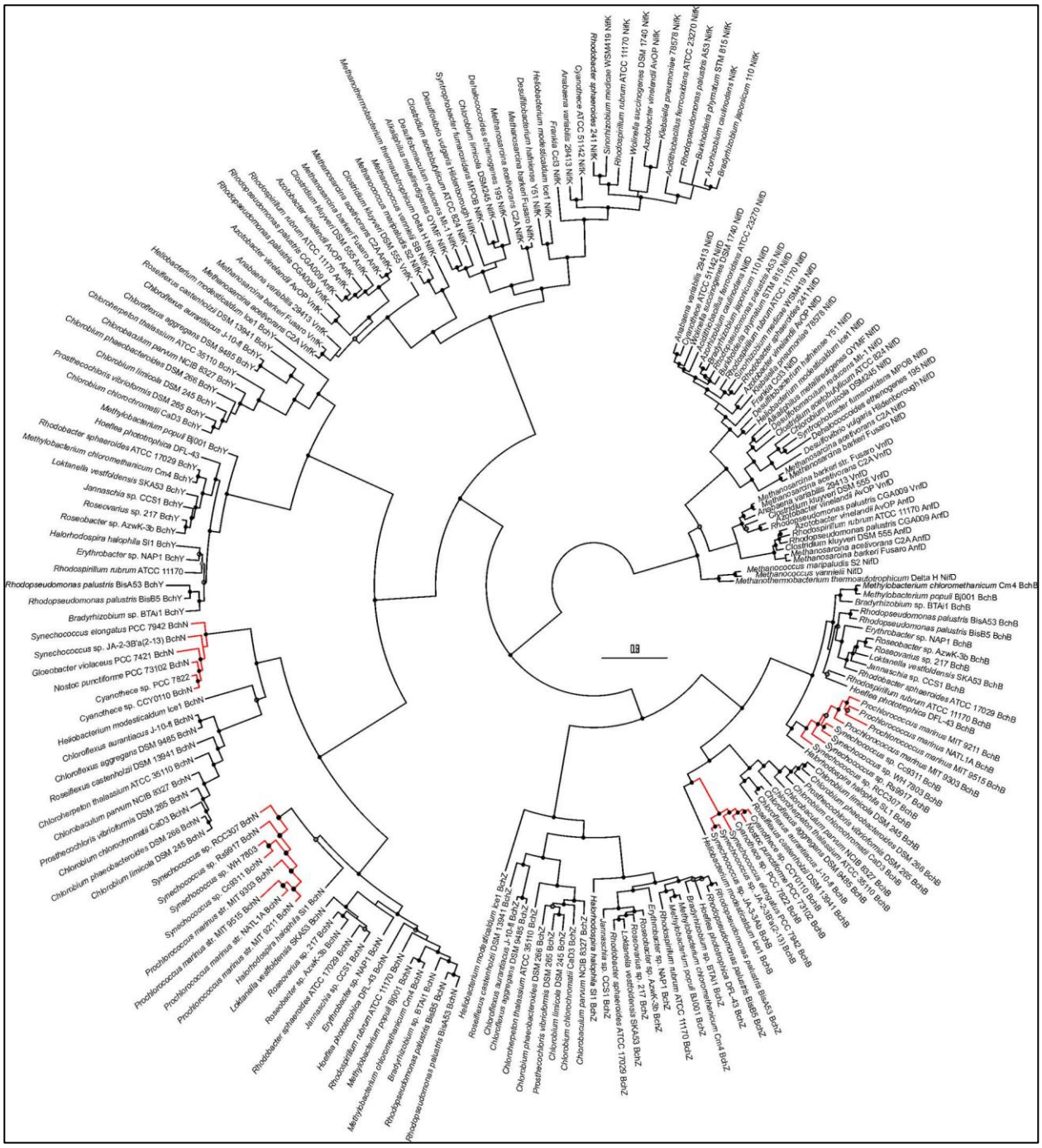


Figure A7. Bayesian consensus phylogenetic tree of Nif/Vnf/AnfDK and BchNBYZ proteins from taxa representing the primary lineages for each locus. Black circles at nodes denote >90% posterior probability (PP), gray circles at nodes denote >80% PP, open circles denote >70% PP, and no symbol denotes 50 to 70% PP. Nodes with <50% PP were collapsed; bar equals 9 substitutions per 10 sequence positions. Red lineages contain deduced amino acid sequences derived from oxygenic phototrophs.

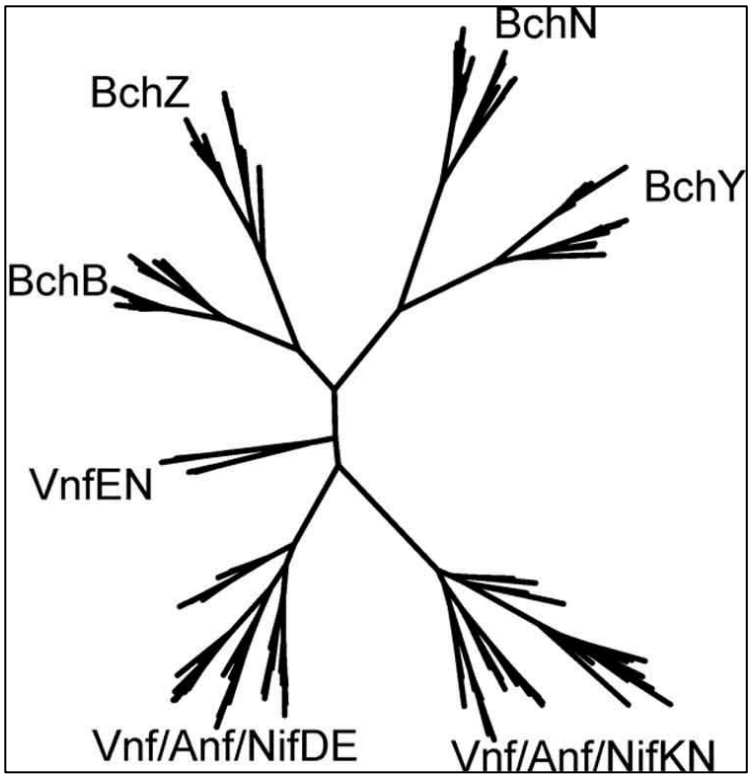


Figure A8. Bayesian consensus phylogenetic tree of representatives of Vnf/Anf/NifDKEN and BchNBYZ loci.

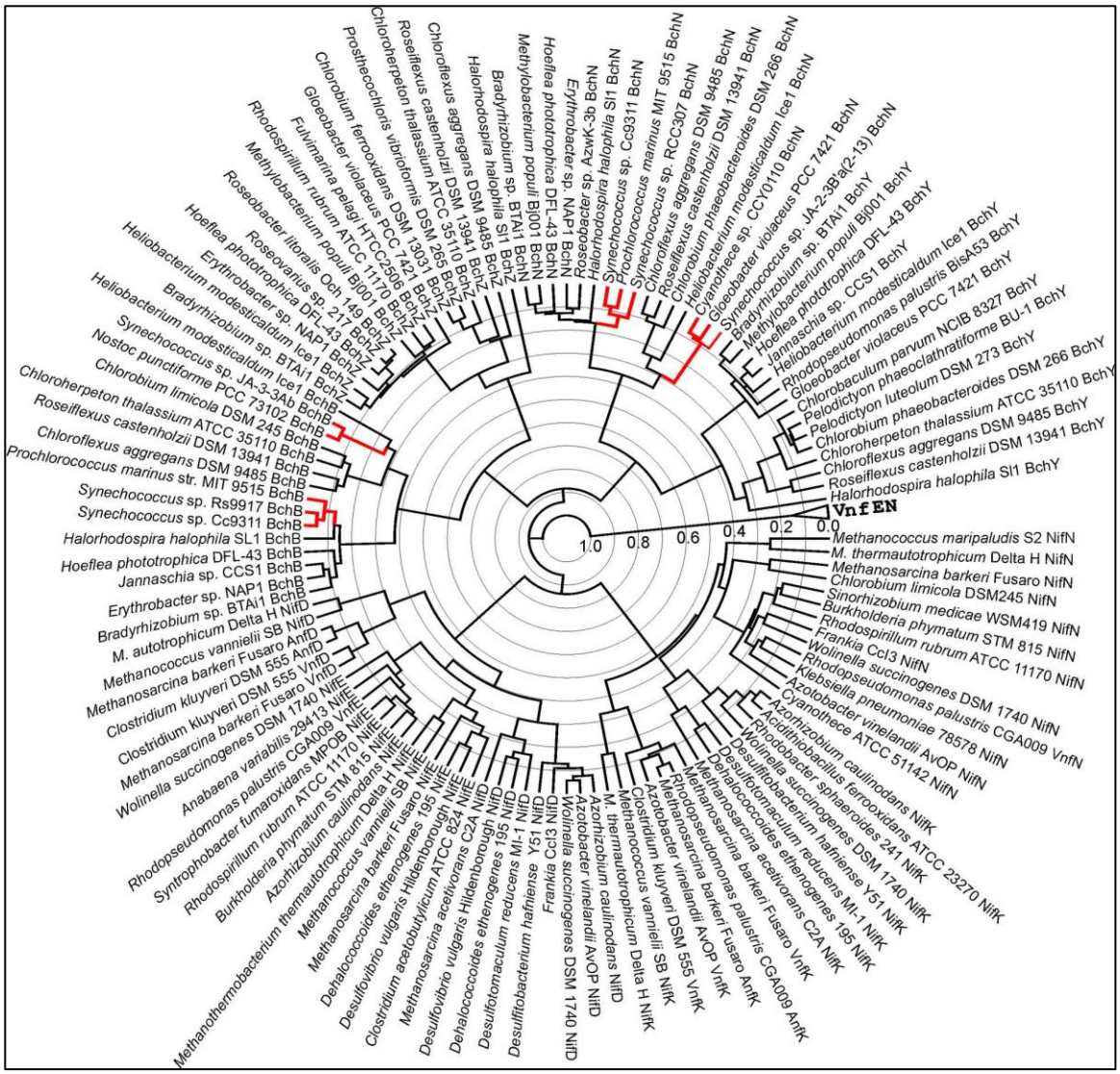


Figure A9. Rate-smoothed radial chronogram of representative Vnf/Anf/NifDKEN and BchNBYZ loci as estimated using the Penalized Likelihood method (*19*). Red lineages denote BchNB deduced amino acid sequences derived from oxygenic phototrophs. The root (VnfEN) age was fixed to 1.0 units.

Literature Cited

1. Boyd, E. S., Anbar, A. D., Miller, S., Hamilton, T. L., Lavin, M., and Peters, J. W. (2011) A late methanogen origin for molybdenum-dependent nitrogenase, *Geobiol. 9*, 221–232.

2. Soboh, B., Boyd, E. S., Zhao, D., Peters, J. W., and Rubio, L. M. (2010) Substrate specificity and evolutionary implications of a NifDK enzyme carrying NifB-co at its active site, *FEBS Letters 584*, 1487-1492.

3. Abascal, F., Zardoya, R., and Posada, D. (2005) ProtTest: selection of best-fit models of protein evolution, *Bioinformatics 2*, 2104-2105.

4. Huelsenbeck, J. P., and Ronquist, R. (2001) MRBAYES: Bayesian inference of phylogeny, *Bioinformatics 17*, 754–755.

5. Guindon, S., and Gascuel, O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood., *Systematic Biology 52*, 696-704.

6. Webb, C. O., Ackerly, D. D., and Kembel, S. W. (2008) Phylocom: software for the analysis of phylogenetic community structure and trait evolution, *Bioinformatics 24*, 2098-2100.

7. Nielsen, M., Lundegaard, C., Lund, O., and Petersen, T. N. (2010) CPHmodels-3.0-remote homology modeling using structure-guided sequence profiles, *Nucleic Acids Research 38*, W576-W581.

8. Georgiadis, M. M., Komiya, H., Chakrabarti, P., Woo, D., Kornuc, J. J., and Rees, D. C. (1992) Crystallographic structure of the nitrogenase iron protein from *Azotobacter vinelandii*, *Science 257*, 1653-1659.

9. Peters, J. W., Stowell, M. H., Soltis, S. M., Finnegan, M. G., Johnson, M. K., and Rees, D. C. (1997) Redox-dependent structural changes in the nitrogenase P-cluster, *Biochemistry 36*, 1181-1187.

10. Chiu, H., Peters, J. W., Lanzilotta, W. N., Ryle, M. J., Seefeldt, L. C., Howard, J. B., and Rees, D. C. (2001) MgATP-Bound and nucleotide-free structures of a nitrogenase protein complex between the Leu 127 Delta-Fe-protein and the MoFe-protein. , *Biochem. 40*, 641-650.

11. Mayer, S. M., Gormal, C. A., Smith, B. E., and Lawson, D. M. (2002) Crystallographic analysis of the MoFe protein of nitrogenase from a nifV mutant of Klebsiella pneumoniae identifies citrate as a ligand to the molybdenum of iron molybdenum cofactor (FeMoco), *J. Biol. Chem. 277*, 35263-35266.

12. Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2007) Clustal W and Clustal X version 2.0, *Bioinformatics 23*, 2947-2948.

13. Peters, J. W., Fisher, K., and Dean, D. R. (1995) Nitrogenase structure and function: a biochemical-genetic perspective, *Annu. Rev. Microbiol. 49*, 335-366.

14. Kim, J., and Rees, D. C. (1992) Crystallographic structure and functional implications of the nitrogenase molybdenum-iron protein from *Azotobacter vinelandii*, *Nature 360*, 553-560.

15. Ronquist, F., and Huelsenbeck, J. P. (2003) MRBAYES 3: Bayesian phylogenetic inference under mixed models, *Bioinformatics 19*, 1572-1574.

16. Huelsenbeck, J. P., and Ronquist., R. (2001) MRBAYES: Bayesian inference of phylogeny, *Bioinformatics 17*, 754-755.

17. Page, R. D. M. (1996) TREEVIEW: An application to display phylogenetic trees on personal computers. , *Comp. Appl. Biosci. 12*, 357-358.

18. Xiong, J., Fischer, W. M., Inoue, K., Nakahara, M., and Bauer, C. E. (2000) Molecular evidence for the early evolution of photosynthesis, *Science 289*.

19. Sanderson, M. J. (2002) Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. , *Mol. Biol. Evol. 19*, 101-109.

20. Staples, C. R., Lahiri, S., Raymond, J., Von Herbulis, L., Mukhophadhyay, B., and Blankenship, R. E. (2007) Expression and association of group IV nitrogenase NifD and NifH homologs in the non-nitrogen-fixing archaeon *Methanocaldococcus jannaschii*, *J Bacteriol. 189*, 7392-7398.

21. Muraki, N., Nomata, J., Ebata, K., Mizoguchi, T., Shiba, T., Tamiaki, H., Kurisu, G., . , and Fujita, Y. (2010) X-ray crystal structure of the light-independent protochlorophyllide reductase, *Nature 465*, 110-114.

22. Raymond, J., Siefert, J. L., Staples, C. R., and Blankenship, R. E. (2004) The natural history of nitrogen fixation, *Mol. Biol. Evol 21*, 541-554.

APPENDIX B:

SUPPORTING INFORMATION FOR CHAPTER 3

Chapter 3: Environmental Constraints Underpin the Distribution and Phylogenetic Diversity of *nifh* in the Yellowstone Geothermal Complex

Published: Microbial Ecology, **2011**, *61*, 860 – 870.

Further Supporting Materials are available free of charge via the Internet at http://www.springerlink.com/content/27j2748v33770203/supplementals/. This contains 1 figure and 2 tables.

Table B1. Pairwise distances[a] between sampling sites.

Site[b]	NL13	IG6	HL11	NG16	IG11	IG5	NL12	NG14	NL4	HL5	NG24	NG21	NG13
NL13	0.00	27.11	52.35	3.21	27.11	27.11	0.32	3.21	<0.01	51.56	2.9	2.55	3.21
IG6		0.00	37.81	25.37	<0.01	<0.01	27.38	25.37	27.11	37.32	25.19	25.98	25.37
HL11			0.00	49.2	37.81	37.81	52.44	49.2	52.35	0.89	49.46	49.93	49.2
NG16				0.00	25.37	25.37	3.25	<0.01	3.21	48.41	0.52	0.74	<0.01
IG11					0.00	<0.01	27.38	25.37	27.11	37.32	25.19	25.98	25.37
IG5						0.00	3.25	3.25	0.32	51.65	2.99	2.56	3.25
NL12							0.00	3.25	0.32	51.65	2.99	2.56	3.25
NG14								0.00	3.21	48.41	0.52	0.74	<0.01
NL4									0.00	51.56	2.9	2.55	3.21
HL5										0.00	48.66	49.13	48.41
NG24											0.00	0.79	0.52
NG21												0.00	0.74
NG13													0.00

[a]Distances measured in kilometers

[b]Geothermal spring designations correspond to those presented in Table 1 of the main text.

Table B2. Clone frequencies, accession numbers, and affiliations in each of the 13 environments studied.

Freq[a]	Site[b]	Closest Affiliated Sequence[c]	Phylum[c]	% Identity (% Similarity) [c]
4	HL5 (8.3%), NG16 (8.0%)	*Thermocrinis albus* DSM 14484	*Aquificae*	91 (96)
1	IG11 (4.7%)	*Thermocrinis albus* DSM 14484	*Aquificae*	90 (95)
1	NG14 (4.7%)	*Thermocrinis albus* DSM 14484	*Aquificae*	93 (97)
2	NL4 (11.1%)	*Anabaena variabilis* ATCC 29413	*Cyanobacteria*	89 (97)
1	IG6 (4.5%)	*Hydrogenobacter thermophilus* TK-6	*Aquificae*	95 (99)
19	HL5 (4.2%), IG11 (4.7%), NG16 (12.0%), NG14 (14.3%), NL13 (71.4%), NL4 (5.5%)	*Burkholderia xenovorans* LB400	*Betaproteobacteria*	99 (100)
7	NG16 (20.0%), NL12 (12.5%)	*Burkholderia xenovorans* LB400	*Betaproteobacteria*	97 (99)
1	NL13 (7.1%)	*Burkholderia xenovorans* LB400	*Betaproteobacteria*	98 (99)
2	NG16 (4.0%), NL12 (6.3%)	*Acidithiobacillus ferrooxidans* ATCC 53993	*Gammaproteobacteria*	100 (100)
9	NG16 (4.0%), NG24 (9.5%), NL12 (6.3%), NL4 (22.2%), NG13 (5.9%)	*Acidithiobacillus ferrooxidans* ATCC 53993	*Gammaproteobacteria*	98 (99)
2	NL4 (11.1%)	*Acidithiobacillus ferrooxidans* ATCC 53993	*Gammaproteobacteria*	98 (99)
3	IG5 (5.9%), NG13 (11.7%)	*Burkholderia tropica* strain Ppe8	*Betaproteobacteria*	94 (96)
23	IG11 (9.5%), NG21 (20.0%), NG24 (14.3%), NL12 (12.5%), NL4 (11.1%), NG13 (52.9%)	*Burkholderia tropica*	*Betaproteobacteria*	95 (96)
1	NG24 (4.6%)	*Burkholderia tropica*	*Betaproteobacteria*	94 (95)
1	NG13 (5.9%)	*Methylacidiphilum fumariolicum*	*Verrucomicrobia*	99 (100)
1	NG13 (5.9%)	*Burkholderia tropica*	*Betaproteobacteria*	95 (96)
2	NG21 (4.0%), NG24 (4.6%)	*Burkholderia tropica*	*Betaproteobacteria*	96 (97)
6	NG21 (12.0%), NG24 (9.5%), NL12 (6.3%)	*Methylocella silvestris* BL2	*Alphaproteobacteria*	95 (97)
1	NG24 (4.6%)	*Methylocella silvestris* BL2	*Alphaproteobacteria*	94 (96)
3	NL4 (16.6%)	*Acidithiobacillus*	*Gammaproteobacteria*	97 (99)

		ferrooxidans ATCC 53993		
1	NG13 (5.9%)	*Burkholderia tropica*	*Betaproteobacteria*	95 (96)
1	NG21 (4.0%)	*Burkholderia tuberum*	*Betaproteobacteria*	93 (99)
2	NG16 (4.0%), NG14 (4.7%)	*Methylacidiphilum fumariolicum*	*Verrucomicrobia*	99 (99)
2	NG14 (9.5%)	*Methylacidiphilum fumariolicum*	*Verrucomicrobia*	100 (100)
4	NG24 (9.5%), NG13 (11.8%)	*Methylacidiphilum fumariolicum*	*Verrucomicrobia*	99 (100)
22	NG21 (56.0%), NG24 (19.0%), NL12 (25.0%)	*Dechloromonas* sp. SIUL	*Betaproteobacteria*	93 (99)
1	NG24 (4.6%)	*Dechloromonas* sp. SIUL	*Betaproteobacteria*	92 (97)
2	NG21 (4.0%), NL12 (6.3%)	*Dechloromonas* sp. SIUL	*Betaproteobacteria*	92 (98)
2	NG24 (4.6%), NL4 (5.5%)	*Ectothiorhodospira mobilis*	*Gammaproteobacteria*	91 (97)
2	NG24 (9.5%)	*Ectothiorhodospira shaposhnikovii* DSM 2111	*Gammaproteobacteria*	92 (98)
2	NG24 (4.6%), NL4 (5.5%)	*Ectothiorhodospira haloalkaliphila*	*Gammaproteobacteria*	98 (100)
1	IG5 (5.9%)	*Dechloromonas* sp. SIUL	*Betaproteobacteria*	98 (100)
2	HL5 (8.3%)	*Mastigocladus laminosus* CCMEE 5201	*Cyanobacteria*	92 (97)
21	HL5 (54.2%), IG11 (4.7%), NG16 (16.0%), IG6 (4.5%), HL11 (12.0%)	*Mastigocladus laminosus* CCMEE 5201	*Cyanobacteria*	93 (98)
1	IG11 (4.7%)	*Mastigocladus laminosus* CCMEE 5201	*Cyanobacteria*	92 (97)
7	IG5 (11.7%), IG11 (14.3), HL11 (8.0%)	*Anabaena variabilis* ATCC 29413	*Cyanobacteria*	96 (98)
1	IG11 (4.7%)	*Anabaena variabilis* ATCC 29413	*Cyanobacteria*	95 (97)
14	HL5 (12.5%), IG11 (33.3%), NG16 (4.0%), NL13 (14.3%), HL11 (4.0%)	*Synechococcus* sp. JA-3-3Ab	*Cyanobacteria*	100 (100)
1	IG11 (4.7%)	*Anabaena variabilis* ATCC 29413	*Cyanobacteria*	90 (97)
1	IG11 (4.7%)	*Synechococcus* sp. JA-3-3Ab	*Cyanobacteria*	99 (99)
1	NL4 (5.5%)	*Thermocrinis albus* DSM 14484	*Aquificae*	92 (96)
2	HL5 (8.3%)	*Synechococcus* sp. JA-3-3Ab	*Cyanobacteria*	99 (99)
1	HL5 (4.2%)	*Fischerella* sp. UTEX	*Cyanobacteria*	90 (97)

		'LB 1931'		
42	IG5 (64.7%), IG11(4.7%), IG6 (81.8%), HL11 (48.0%)	*Mastigocladus laminosus* CCMEE 5201	*Cyanobacteria*	98 (99)
1	IG6 (4.5%)	*Mastigocladus laminosus* CCMEE 5201	*Cyanobacteria*	98 (99)
1	HL11 (4.0%)	Mastigocladus laminosus CCMEE 5186	*Cyanobacteria*	98 (99)
1	HL11 (4.0%)	*Mastigocladus laminosus* CCMEE 5201	*Cyanobacteria*	99 (100)
1	IG6 (4.5%)	*Mastigocladus laminosus* CCMEE 5201	*Cyanobacteria*	99 (100)
1	HL11 (4.0%)	*Anabaena variabilis* ATCC 29413	*Cyanobacteria*	96 (98)
1	IG5 (5.9%)	*Fischerella* sp. UTEX 'LB 1931'	*Cyanobacteria*	98 (99)
1	HL11 (4.0%)	*Mastigocladus laminosus* CCMEE 5201	*Cyanobacteria*	99 (100)
1	IG11 (4.7%)	*Leptolyngbya boryana* IAM M-101	*Cyanobacteria*	99 (100)
1	NL4 (5.5%)	*Acidithiobacillus ferrooxidans* ATCC 53993	*Gammaproteobacteria*	98 (99)
1	HL11 (4.0%)	*Thermocrinis albus* DSM 14484	*Aquificae*	97 (98)
3	NG16 (4.0%), NG14 (9.5%)	*Mastigocladus laminosus* CCMEE 5201	*Cyanobacteria*	100 (100)
2	NG14 (9.5%)	*Mastigocladus laminosus* CCMEE 5201	*Cyanobacteria*	84 (96)
6	NG14 (25.6%)	*Mastigocladus laminosus* CCMEE 5201	*Cyanobacteria*	85 (96)
1	HL11 (4.0%)	*Synechococcus* sp. JA-2-3B'a(2-13)	*Cyanobacteria*	100 (100)
1	IG5 (5.9%)	*Geobacter* sp. M21	*Deltaproteobacteria*	96 (97)
9	NG16 (16.0%), NG14 (9.5%), NL13(21.4%)	*Syntrophobacter fumaroxidans* MPOB	*Deltaproteobacteria*	94 (98)
2	NG16 (4.0%), NL12 (6.3%)	*Syntrophobacter fumaroxidans* MPOB	*Deltaproteobacteria*	94 (97)
3	NG16 (4.0%), NG14 (9.5%)	*Syntrophobacter fumaroxidans* MPOB	*Deltaproteobacteria*	93 (97)
1	NG16 (4.0%)	*Syntrophobacter fumaroxidans* MPOB	*Deltaproteobacteria*	92 (96)
2	HL11 (8.0%)	*Opitutaceae bacterium* TAV2	*Verrucomicrobia*	94 (95)

[a]The frequency of the clone recovered in the present study.
[b]The relative abundance of the clone in each environment that it was identified

[c]Closest cultured representative, phylum, order, and % sequence identity and similarity as determined using BLASTp.

Table B3. GPS coordinates for the YNP thermal features where *nifH* clone libraries were constructed and sequenced.

Site[a] (YNP ID)	GPS[b]	
NL13 (NA)	N,44.751917	W,110.728556
IG6 (NA)	N,44.531833	W,110.875917
HL11 (NA)	N,44.304361	W,110.522833
NG16 (NBB113)	N,44.726611	W,110.709139
IG11 (NA)	N,44.531833	W,110.875917
IG5 (NA)	N,44.531833	W,110.875917
NL12 (NMCNN030)	N,44.753528	W,110.725222
NG14 (NBB113)	N,44.726611	W,110.709139
NL4 (NA)	N,44.751917	W,110.728556
HL5 (HLFNN164)	N,44.312250	W,110.521306
NG24 (NRHA014)	N,44.727500	W,110.715639
NG21 (NHSP101)	N,44.733278	W,110.709778
NG13 (NBB13)	N,44.726611	W,110.709139

[a]NL, Nymph Lake; IG, Imperial Geyser within the Lower Geyser Basin; HL, Heart Lake Geyser Basin; NG, Norris Geyser Basin

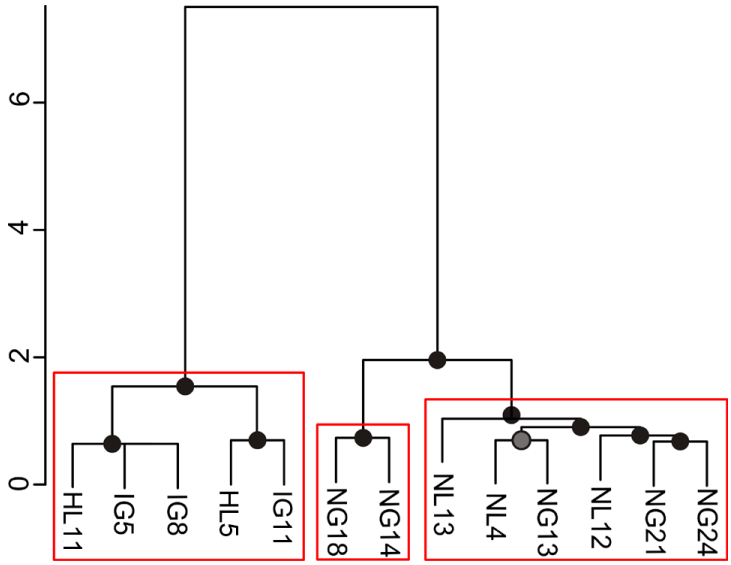


Figure B1. Hierarchical clustering diagram of the relationships in D_p distance based on the Paup smoothed tree, using 1000 permutations and the Wards method of clustering. The values at the nodes are the average unbiased P-*values* and the red boxes denote clusters with a *P*-value <0.05. Sampling site abbreviations correspond to those presented in Table 1.

APPENDIX C:

SUPPORTING INFORMATION FOR CHAPTER 4

Chapter 4: Biological Nitrogen Fixation in Acidic High Temperature Geothermal Springs in Yellowstone National Park, Wyoming

Published: *Environmental Microbiology*, **2011**, *13*, 2204 – 2215

Table C1. Clone frequencies, accession numbers, and affiliations in each of the environments studied.

This Study	GenBank (ID)	Frequency	Closest Blast Hit (tBLASTn)	Identities (%)
YNP-5	NGB1-9	19	Uncultured organism from YNP (GQ426268)	107/107 (100)
YNP-1	NGB1-24	7	Uncultured organism from YNP (HM113543)	107/107 (100)
YNP-10	NGB2-8	12	Uncultured organism from YNP (GQ426266)	106/107 (99)
YNP-3	NGB2-12	11	Uncultured organism from YNP (GQ426268)	106/107 (99)
YNP-2	NGB3-2	7	Uncultured organism from YNP (HM113541)	107/107 (100)
YNP-6	NGB3-3	11	Uncultured organism from YNP (GQ426271)	105/107 (99)
YNP-4	NGB4-2	7	Uncultured organism from YNP (GQ426268)	106/107 (99)
YNP-9	NL2-4	10	Uncultured organism from YNP (GQ426268)	107/107 (100)
YNP-11	NGB4-4	10	Uncultured organism from YNP (GQ426271)	105/107 (99)
YNP-7	NL1-11	9	Uncultured organism from YNP (GQ426268)	106/107 (99)
YNP-8	NL1-6	9	Burkholderia xenovorans LB400	106/107 (99)
NG4 HFS	NG4 HFS	1	Uncultured organism from YNP (GQ426268)	107/107 (100)

Table C2. GPS coordinates for the YNP thermal features where *nifH* clone libraries were constructed and sequenced.

Site[a]	GPS (N,W)
NL5	44.7519, -110.7286
NG23	44.7318, -110.7125
NG19	44.7324, -110.7098
NG1	44.7293, -110.7125
NL1[e]	44.7519, -110.7286
NG18	44.7279, -110.7103
NL12	44.7535, -110.7252
NG2	44.7328, -110.7096
NL9	44.7519, -110.7286
NL2	44.7519, -110.7286
NG3[e]	44.7275, -110.7156
NG4[e]	44.7332, -110.7098
NG12	44.7286, -110.7126
NG22	44.7332, -110.7098

[a]NL, Nymph Lake; IG, Imperial Geyser within the Lower Geyser Basin; HL, Heart Lake Geyser Basin; NG, Norris Geyser Basin

APPENDIX D:

SUPPORTING INFORMATION FOR CHAPTER 5

Chapter 5: Transcriptional Profiling of Nitrogen Fixation in *Azotobacter Vinelandii*

Published: Journal of Bacteriology, **2011**, *193*, 4477 – 4486

Further Supporting Information is available free of charge via the Internet at http://jb.asm.org. This information includes supplemental tables of the changes in transcript levels of *nif*-specific genes, genes implicated in the global response for Mo-dependent nitrogen fixation, *nif*-, *vnf*-, and *anf*-specific genes, σ^{54}-dependent activators, *rnf*, and *rnf1* operons and the *fix* operon, the changes in transcript abundance of the top 10 up- and down-regulated genes under each growth conditions, the 15 most abundant mRNA species under each condition, transcript counts for differentially expressed genes chosen for RT-PCR and qRT-PCR analyses and a one figure of the results of qRT-PCR analysis.

Table D1. Changes in transcript levels of the *nif*-specific genes necessary for Mo-dependent nitrogen fixation. (FC, Fold Change; Nif, Mo-dependent nitrogenase)

Major *nif* cluster			
Locus Tag	FC Nif/Control	Gene name	Gene Product
Avin_01360	24.5	*orf13*	Conserved hypothetical protein
Avin_01370	29.5	*orf12*	Conserved hypothetical protein
Avin_01380	143.8	*nifH*	Nitrogenase iron protein, NifH
Avin_01390	53.1	*nifD*	Nitrogenase molybdenum-iron protein alpha chain, NifD
Avin_01400	55.1	*nifK*	Nitrogenase molybdenum-iron protein beta chain, NifK
Avin_01410	1.9	*nifT*	Nitrogen fixation protein
Avin_01420	8.1	*nifY*	Nitrogenase iron molybdenum cofactor biosynthesis protein
Avin_01430	4.0	*orf1*	Conserved hypothetical protein
Avin_01440	13.1	*lrv*	Nitrogen fixing leucine rich variant repeat 4Fe-4S cluster protein
Avin_01450	6.3	*nifE*	Nitrogenase MoFe cofactor biosynthesis protein, NifE
Avin_01470	3.8	*nifN*	Nitrogenase MoFe cofactor biosynthesis protein, NifN
Avin_01480	2.0	*nifX*	Nitrogenase MoFe cofactor biosynthesis protein, NifX
Avin_01490	2.0	*orf3*	Conserved hypothetical protein
Avin_01500	4.6	*orf4*	Conserved hypothetical protein
Avin_01510	5.2	-	Nitrogen fixation (4Fe-4S) ferredoxin-like protein
Avin_01520	2.3	*fes_I*	Nitrogen fixation (2Fe-2S) ferredoxin (Shethna I protein), FeS1
Avin_01530	7.5	-	Conserved hypothetical protein
Avin_01540	3.4	-	Conserved hypothetical protein
Avin_01550	11.0	-	Conserved hypothetical protein
Avin_01560	6.7	-	Conserved hypothetical protein
Avin_01570	6.7	*orf11*	Conserved hypothetical protein
Avin_01580	1.3	*orf12*	ABC transporter, ATP-binding protein
Avin_01600	3.5	*orf5*	Conserved hypothetical protein
Avin_01610	23.7	*iscAnif*	Nitrogen fixation Fe-S cluster assembly protein, IscAnif
Avin_01620	13.1	*nifU*	Nitrogen fixation Fe-S cluster scaffold protein, NifU
Avin_01630	8.0	*nifS*	Nitrogen fixation cysteine desulfurase, NifS
Avin_01640	6.7	*nifV*	Nitrogen fixation homocitrate synthase, NifV
Avin_01650	6.3	*cysE1nif*	Nitrogen fixation serine O-acetyltransferase, CysE1
Avin_01660	9.2	*orf8*	Conserved hypothetical protein
Avin_01670	4.9	*nifW*	Nitrogen fixation protein, NifW
Avin_01680	4.9	*nifZ*	Nitrogen fixation protein, NifZ
Avin_01690	6.8	*nifM*	Nitrogen fixation cis-trans peptidyl prolyl isomerase, NifM
Avin_01700	5.5	*clpXnif*	Nitrogen fixation protein ORF9,

			ClpX
Avin_01710	8.3	*nifF*	Flavodoxin, NifF
Minor *nif* cluster			
Locus Tag	FC Nif/Control	Gene name	Gene Product
Avin_50990	0.3	*nifL*	Nitrogen fixation regulatory protein, NifL
Avin_51000	0.9	*nifA*	Nif-specific σ^{54}-dependent transcriptional activator protein, NifA
Avin_51010	18.0	*nifB*	Nitrogenase cofactor biosynthesis protein, NifB
Avin_51020	16.8	*fdx*	Ferredoxin protein
Avin_51030	16.4	*nifO*	Nitrogenase-associated protein, NIfO
Avin_51040	12.0	*nifQ*	Nitrogen fixation cofactor assembly protein, NifQ
Avin_51050	5.6	*rhd*	Rhodanese/sulfurtransferase-like protein
Avin_51060	4.1	*grx5*	Glutaredoxin-related protein

Table D2. Changes in transcript levels of genes implicated in the global respones *of A. vinelandii* for Mo-dependent nitrogen fixation. (FC, Fold Change; Nif, Mo-dependent nitrogenase)

Molybdate Transport			
Locus Tag	FC Nif/Control	Gene name	Gene Product
Avin_01280	1.9	-	ABC transporter, membrane spanning protein
Avin_01290	1.1	-	ABC transporter, membrane spanning protein
Avin_01300	1.7	*modA2*	Periplasmic molybdate-binding protein, ModA2
Avin_50650	4.0	*modC1*	Mo transporter, inner membrane ATP-binding component, ModC1
Avin_50660	6.3	*modB1*	Mo transporter membrane protein, ModB1
Avin_50670	5.2	*modA1*	Molybdenum transporter, periplasmic molybdate-binding protein
Avin_50680	3.4	*modE*	Mo regulation, Mo processing homeostasis
Avin_50690	3.2	*modG*	Mo processing, homeostasis
Avin_50700	5.3	-	ABC transporter ATP binding component
Avin_50710	1.0	-	ABC type tungstate transporter, permease component
Avin_50720	1.5	-	ABC transporter, permease protein
Avin_50730	1.2	*modA3*	Molybdate ABC transporter, periplasmic molybdate-binding protein

Pilus			
Locus Tag	FC Nif/Control	Gene name	Gene Product
Avin_03090	60.3	*pilT*	Pilus retraction protein
Avin_03180	105.2	*pilG*	Type IV pili response regulator, PilG
Avin_11820	26.8	*flmT*	Type IV pilus biogenesis protein, FlmT
Avin_11830	138.5	*pilE*	Type IV PilE-like pilus assembly protein
Avin_11840	94.4	*pilV*	Type IV pilus assembly transmembrane protein, PilV-like protein
Avin_11850	24.7	*pilW*	Type IV pilus assembly transmembrane protein, PilW-like protein
Avin_11870	19.3	*pilY1*	Type IV pilus assembly protein, PilY1-like protein
Avin_12070	13.4	*pilD*	Type IV Pilus Prepilin peptidase, PilD
Avin_12080	4.9	*pilC*	Type IV pilus biogenesis protein
Avin_12090	25.2	*pilB*	Type IV pilus assembly protein
Avin_12104	97.9	*pilA*	Fimbrial protein
Avin_45260	17.2	*pilQ*	Bacterial secretion system protein
Avin_45270	66.6	*pilP*	Pilus assembly protein
Avin_45280	18.6	*pilO*	Pilus assembly protein
Avin_45290	90.7	*pilN*	Fimbrial assembly protein
Avin_45300	214.2	*pilM*	Type IV pilus assembly protein

ISC Machinery			
Locus Tag	FC Nif/Control	Gene name	Gene product
Avin_40340	0.0	*iscX*	Conserved hypothetical protein
Avin_40350	0.0	*fdx*	Isc ferredoxin
Avin_40360	0.0	*hscA*	Fe-S protein assembly chaperone, HscA
Avin_40370	0.0	*hscB*	Co-chaperone Hsc20
Avin_40380	0.0	*iscA*	Iron-sulfur biogenesis protein
Avin_40390	0.0	*iscU*	Iron-sulfur cluster assembly scaffold protein

Avin_40400	0.0	*iscS*	Cysteine desulfurase IscS
Avin_40410	0.0	*iscR*	Iron-sulphur cluster assembly transcription factor, IscR

Fe-S Cluster Biosynthesis			
Locus Tag	FC Nif/Control	Gene name	Gene product
Avin_28760	0.0	*nfuA*	NfuA protein
Avin_46030	0.0	*erpA*	Essential respiratory protein A, ErpA

Ferrous Iron Transport			
Locus Tag	FC Nif/Control	Gene name	Gene product
Avin_22720	2.4	*feoA*	Ferrous iron transport protein A
Avin_22730	2.0	*feoB*	Ferrous iron transport protein B
Avin_22740	2.0	*feoC*	Ferrous iron transport protein, feoC-like protein

Fe Siderophore Machinery			
Locus Tag	FC Nif/Control	Gene name	Gene product
Avin_09310	0.0	-	Siderophore biosynthesis protein, IucA/IucC family
Avin_09320	0.0	-	Transporter, MFS superfamily
Avin_09330	0.0	-	Siderophore biosynthesis protein, IucA/IucC family
Avin_21180	0.0	*entA*	2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase
Avin_21190	0.1	*entF*	Enterochelin sythetase component F
Avin_21200	0.0	*entB*	Isochorismatase (2,3 dihydro-2,3 dihydroxybenzoate synthase)
Avin_21210	0.0	*entE*	Enterobactin synthetase component E (2,3-dihydroxybenzoate-AMP ligase)
Avin_21220	0.0	*csbC*	Isochorismate synthase
Avin_21230	0.0	*csbX*	Catecholate siderophore efflux pump, MFS_1 family

[NiFe]-Hydrogenase			
Locus Tag	FC Nif/Control	Gene name	Gene product
Avin_04360	2.1	-	Soluble hydrogenase, alpha or beta chain
Avin_04370	3.7	-	Cyclic nucleotide-binding protein, hydrogenase accessory protein, HoxI
Avin_04380	2.2	-	Soluble hydrogenase gamma subunit
Avin_04390	1.8	*hoxY*	Soluble hydrogenase delta subunit, HoxY
Avin_04400	0.8	*hoxH*	Soluble nickel-dependent hydrogenase, large subunit, HoxH
Avin_04410	0.4	*hoxW*	Soluble hydrogenase processing peptidase, HoxW
Avin_50440	2.7	*hypE*	Hydrogenase expression/formation protein, HypE
Avin_50450	1.4	*hypD*	Hydrogenase expression/formation protein, HypD
Avin_50460	1.6	*hypC*	Hydrogenase assembly chaperone, HypC/HupF
Avin_50470	1.3	*hypF*	Hydrogenase maturation carbamoyltransferase, HypF
Avin_50480	3.2	*hypB*	Hydrogenase nickel incorporation protein, HypB
Avin_50490	15.5	*hypA*	Hydrogenase nickel incorporation protein, HypA
Avin_50500	4.6	*hoxV*	Hydrogenase expression/formation protein, HoxV
Avin_50510	8.1	*hoxT*	Hydrogenase expression/formation protein, HoxT
Avin_50520	7.8	*hoxR*	Rubredoxin-type Fe(Cys)4 protein, HoxR
Avin_50530	3.6	*hoxQ*	Hydrogenase expression/formation protein, HoxQ

Avin_50540	2.0	*hoxO*	Hydrogenase expression/formation protein, HoxO
Avin_50550	1.7	*hoxL*	Hydrogenase assembly chaperone, HoxL (HypC/HupF family)
Avin_50560	2.5	*hoxM*	Hydrogenase expression/formation protein, HoxM
Avin_50570	7.6	*hoxZ*	Ni/Fe-hydrogenase, b-type cytochrome subunit, HoxZ
Avin_50580	4.5	*hoxG*	Membrane bound nickel-dependent hydrogenase, large subunit, HoxG
Avin_50590	5.0	*hoxK*	Uptake hydrogenase small subunit (Precursor), HoxK

Terminal Oxidases

Locus Tag	FC Nif/Control	Gene name	Gene product
Avin_00950	0.4	-	Cytochrome oxidase assembly protein
Avin_00960	0.2	-	Conserved hypothetical protein
Avin_00970	0.4	-	Surfeit locus 1-like protein
Avin_00980	1.1	-	Conserved hypothetical protein
Avin_00990	0.2	-	Cytochrome C Oxidase, Subunit III
Avin_01000	0.3	-	Cytochrome c oxidase assembly protein CtaG/Cox11
Avin_01010	0.4	*ctaD*	Cytochrome c oxidase, subunit I
Avin_01020	0.8	-	Cytochrome c oxidase, subunit II
Avin_11040	0.4	-	Cytochrome bd ubiquinol oxidase, subunit II
Avin_11050	0.7	-	Cytochrome bd ubiquinol oxidase, subunit I
Avin_11170	4.5	*coxB*	Cytochrome o ubiquinol oxidase, subunit II
Avin_11180	6.6	*coxA*	Cytochrome o ubiquinol oxidase, subunit I
Avin_13060	1.1	*petA*	Ubiquinol-cytochrome c reductase, iron-sulfur subunit
Avin_13070	0.8	*petB*	Cytochrome b/b6
Avin_13080	1.9	*petC*	Ubiquinol--cytochrome c reductase, cytochrome c1
Avin_19880	3.2	*cydB*	Cytochrome bd ubiquinol oxidase, subunit II
Avin_19890	3.5	*cydA*	Cytochrome bd ubiquinol oxidase, subunit I
Avin_19910	1.4	*cydR*	Fnr-like negative transcriptional regulator of CydAB
Avin_19940	1.2	*ccoS*	cbb3-type cytochrome oxidase maturation protein
Avin_19950	1.6	*ccoI*	Copper-translocating P-type ATPase
Avin_19960	2.1	*ccoH*	Cytochrome c oxidase accessory protein, CcoH
Avin_19970	1.2	*ccoG*	Cytochrome c oxidase accessory protein, CcoG
Avin_19980	1.1	*ccoP*	Cytochrome c oxidase, cbb3-type, subunit III
Avin_19990	1.6	*ccoQ*	Cytochrome c oxidase, cbb3-type, subunit IV
Avin_20000	1.1	*ccoO*	Cytochrome c oxidase, cbb3-type, subunit II
Avin_20010	1.1	*ccoN*	Cytochrome c oxidase, cbb3-type, subunit I

NADH-Ubiquinone Oxidoreductases

Locus Tag	FC Nif/Control	Gene Name	Gene product
Avin_12000	3.9	*ndh*	Uncoupled NADH:quinone oxidoreductase
Avin_14590	1.3	*nqrA*	Na-translocating NADH-ubiquinone oxidoreductase protein, subunit A
Avin_14600	1.1	*nqrB*	NADH:ubiquinone oxidoreductase, Na(+)-translocating protein, subunit B
Avin_14610	1.1	*nqrC*	NADH:ubiquinone oxidoreductase, Na(+)-translocating protein, subunit C
Avin_14620	1.4	*nqrD*	NADH:ubiquinone oxidoreductase, Na(+)-translocating protein, subunit D

Avin_14630	1.9	*nqrE*	NADH:ubiquinone oxidoreductase, Na(+)-translocating protein subunit E
Avin_14640	1.4	*nqrF*	NADH:ubiquinone oxidoreductase, Na(+)-translocating protein, subunit F
Avin_19530	3.7	*shaAB*	Sodium hydrogen antiporter subunitAB, ShaAB
Avin_19540	1.4	*shaC*	Sodium hydrogen antiporter subunitC, ShaC
Avin_19550	2.8	*shaD*	Sodium hydrogen antiporter subunitD, ShaD
Avin_19560	1.1	*shaE*	Sodium hydrogen antiporter subunitE, ShaE
Avin_19570	1.8	*shaF*	Sodium hydrogen antiporter subunitF, ShaF
Avin_19580	2.8	*shaG*	Sodium hydrogen antiporter subunitG, ShaG
Avin_28440	1.8	*nuoA*	NADH-ubiquinone oxidoreductase, chain alpha
Avin_28450	3.1	*nuoB*	NADH-ubiquinone oxidoreductase, chain beta
Avin_28460	1.0	*nuocd*	NADH-ubiquinone oxidoreductase, chain cd
Avin_28470	1.6	*nuoE*	NADH-ubiquinone oxidoreductase, chain E
Avin_28480	1.4	*nuoF*	NADH-quinone oxidoreductase, F subunit
Avin_28490	0.8	*nuoG*	NADH-quinone oxidoreductase, chain G
Avin_28500	1.3	*nuoH*	Respiratory-chain NADH dehydrogenase, subunit 1, NuoH
Avin_28510	1.8	*nuoI*	NADH-quinone oxidoreductase, chain I
Avin_28520	1.4	*nuoJ*	NADH-ubiquinone/plastoquinone oxidoreductase, chain J
Avin_28530	1.6	*nuoK*	NADH-ubiquinone oxidoreductase, chain K
Avin_28540	1.4	*nuoL*	NADH-plastoquinone oxidoreductase, chain L
Avin_28550	1.1	*nuoM*	Proton-translocating NADH-quinone oxidoreductase, chain M
Avin_28560	1.5	*nuoN*	Proton-translocating NADH-quinone oxidoreductase, chain N

ATP Synthases

Locus Tag	FC Nif/Control	Gene name	
Avin_19670	2.5	-	ATP synthase F1, beta subunit
Avin_19680	1.4	-	H+-transporting two-sector ATPase, delta/epsilon subunit
Avin_19690	1.8	-	H(+)-transporting ATP synthase, gene 1
Avin_19700	2.0	-	lipoprotein
Avin_19710	1.3	-	H+-transporting two-sector ATPase, A subunit
Avin_19720	1.4	-	ATP synthase F0, C subunit
Avin_19730	1.5	-	H+-transporting two-sector ATPase, B/B subunit
Avin_19740	1.6	-	ATP synthase F1, alpha subunit
Avin_19750	2.0	-	H+-transporting two-sector ATPase, gamma subunit
Avin_52150	1.0	*atpC*	F1 sector of membrane-bound ATP synthase, epsilon subunit
Avin_52160	1.6	*atpD*	ATP synthase F1, beta subunit
Avin_52170	1.1	*atpG*	F1 sector of membrane-bound ATP synthase, gamma subunit
Avin_52180	1.4	*atpA*	ATP synthase F1, alpha subunit
Avin_52190	1.4	*atpH*	F1 sector of membrane-bound ATP synthase, delta subunit
Avin_52200	1.6	*atpF*	F0 sector of membrane-bound ATP synthase, subunit B
Avin_52210	1.5	*atpE*	F0 sector of membrane-bound ATP synthase, subunit C
Avin_52220	1.4	*atpB*	F0 sector of membrane-bound ATP synthase, subunit A

Avin_52230	5.4	*atpI*	ATP synthase I chain

Table D3. Changes in transcript levels of the *nif*-, *vnf*-, and *anf*-specific genes necessary for Mo-independent diazotrophy. (FC, Fold Change; Vnf, V-dependent nitrogenase; Anf, Fe-only nitrogenase)

Major *nif* cluster

Locus Tag	FC Vnf/Control	FC Anf/Control	Gene name	Gene product
Avin_01360	3.8	2.9	*orf13*	Conserved hypothetical protein
Avin_01370	7.0	7.6	*orf12*	Conserved hypothetical protein
Avin_01380	2.5	2.6	*nifH*	Nitrogenase iron protein, NifH
Avin_01390	1.6	2.4	*nifD*	Nitrogenase molybdenum-iron protein alpha chain, NifD
Avin_01400	2.2	1.8	*nifK*	Nitrogenase molybdenum-iron protein beta chain, NifK
Avin_01410	0.4	0.5	*nifT*	Nitrogen fixation protein
Avin_01420	1.4	1.8	*nifY*	Nitrogenase iron molybdenum cofactor biosynthesis protein
Avin_01430	0.6	0.6	*orf1*	Conserved hypothetical protein
Avin_01440	2.8	3.2	*lrv*	Nitrogen fixing leucine rich variant repeat 4Fe-4S cluster protein
Avin_01450	2.1	2.6	*nifE*	Nitrogenase MoFe cofactor biosynthesis protein, NifE
Avin_01470	1.3	1.4	*nifN*	Nitrogenase MoFe cofactor biosynthesis protein, NifN
Avin_01480	0.7	1.0	*nifX*	Nitrogenase MoFe cofactor biosynthesis protein, NifX
Avin_01490	0.9	0.9	*orf3*	Conserved hypothetical protein
Avin_01500	1.7	2.2	*orf4*	Conserved hypothetical protein
Avin_01510	2.1	2.3	-	Nitrogen fixation (4Fe-4S) ferredoxin-like protein
Avin_01520	1.1	1.2	*fes_I*	Nitrogen fixation (2Fe-2S) ferredoxin (Shethna I protein), FeS1
Avin_01530	3.0	2.3	-	Conserved hypothetical protein
Avin_01540	1.6	1.7	-	Conserved hypothetical protein
Avin_01550	4.7	6.6	-	Conserved hypothetical protein
Avin_01560	2.8	2.2	-	Conserved hypothetical protein
Avin_01570	3.0	2.7	*orf10*	Conserved hypothetical protein
Avin_01580	1.7	2.3	*orf11*	ABC transporter, ATP-binding protein
Avin_01600	8.3	9.7	*orf5*	Conserved hypothetical protein
Avin_01610	16.0	11.8	*iscAnif*	Nitrogen fixation Fe-S cluster assembly protein, IscAnif
Avin_01620	9.6	7.3	*nifU*	Nitrogen fixation Fe-S cluster scaffold protein, NifU
Avin_01630	7.4	7.0	*nifS*	Nitrogen fixation cysteine desulfurase, NifS
Avin_01640	6.3	5.9	*nifV*	Nitrogen fixation homocitrate synthase, NifV
Avin_01650	3.9	6.1	*cysE1nif*	Nitrogen fixation serine O-acetyltransferase, CysE1
Avin_01660	6.7	8.5	*orf8*	Conserved hypothetical protein
Avin_01670	3.8	3.8	*nifW*	Nitrogen fixation protein, NifW
Avin_01680	3.7	3.7	*nifZ*	Nitrogen fixation protein, NifZ
Avin_01690	4.9	4.9	*nifM*	Nitrogen fixation cis-trans peptidyl prolyl isomerase, NifM
Avin_01700	4.1	3.7	*clpXnif*	Nitrogen fixation protein ORF9, ClpX
Avin_01710	4.4	3.2	*nifF*	Flavodoxin, NifF

Minor *nif* cluster				
Locus Tag	FC Vnf/Control	FC Anf/Control	Gene name	Gene product
Avin_50990	0.6	0.8	*nifL*	Nitrogen fixation regulatory protein, NifL
Avin_51000	1.0	1.1	*nifA*	Nif-specific σ^{54}-dependent transcriptional activator protein, NifA
Avin_51010	7.3	7.9	*nifB*	Nitrogenase cofactor biosynthesis protein, NifB
Avin_51020	8.1	12.0	*fdx*	Ferredoxin protein
Avin_51030	7.7	10.5	*nifO*	Nitrogenase-associated protein, NIfO
Avin_51040	4.9	4.2	*nifQ*	Nitrogen fixation cofactor assembly protein, NifQ
Avin_51050	2.2	1.8	*rhd*	Rhodanese/sulfurtransferase-like protein
Avin_51060	2.3	2.3	*grx5*	Glutaredoxin-related protein

vnf cluster				
Locus Tag	FC Vnf/Control	FC Anf/Control	Gene name	Gene product
Avin_02570	147.5	117.6	*vnfY*	Nitrogen fixation-related protein, VnfY
Avin_02580	33.6	34.7	-	Hypothetical protein
Avin_02590	136.8	3.3	*vnfK*	Vanadium nitrogenase beta subunit, VnfK
Avin_02600	67.9	2.6	*vnfG*	Vanadium nitrogenase, delta subunit, VnfG
Avin_02610	100.8	2.3	*vnfD*	Vanadium nitrogenase, alpha subunit,VnfD
Avin_02650	47.3	102.5	*vnfF*	Vanadium nitrogenase ferredoxin, VnfF
Avin_02660	89.7	17.4	*vnfH*	Vanadium nitrogenase iron protein, VnfH
Avin_02740	17.0	14.8	*vnfX*	Nitrogenase biosynthetic protein, VnfX
Avin_02750	25.5	26.0	*vnfN*	Nitrogenase vanadiun-iron cofactor biosynthesis protein, VnfN
Avin_02770	56.0	54.1	*vnfE*	Nitrogenase vanadium iron cofactor biosynthesis protein, VnfE
Avin_02780	2.9	2.7	*vnfA*	Vanadium nitrogenase σ^{54}-dependent transcriptional activator, VnfA
Avin_02790	2.0	2.2	*vnfU*	NifU C-terminal domain-containing protein, VnfU

anf cluster				
Locus Tag	FC Vnf/Control	FC Anf/Control	Gene name	Gene product
Avin_48950	0.10	7.3	*anfR*	Fe-only nitrogenase accessory protein AnfR
Avin_48960	0.07	10.3	*anfO*	Fe-only nitrogenase accessory protein AnfO
Avin_48970	0.05	30.4	*anfK*	Fe-only nitrogenase, beta subunit, AnfK
Avin_48980	0.05	18.0	*anfG*	Fe-only nitrogenase, delta subunit,

				AnfG
Avin_48990	0.04	38.2	*anfD*	Nitrogenase iron-iron protein, alpha chain, AnfD
Avin_49000	0.02	144.6	*anfH*	Nitrogenase iron protein, AnfH
Avin_49020	2.2	4.6	*anfA*	σ^{54}-dependent transcriptional activator for the iron only nitrogenase, AnfA
Avin_49030	0.6	5.7	*anfU*	NifU C-terminal domain-containing protein, AnfU

Table D4. Changes in transcript levels of the σ^{54}-dependent activators, the *rnf* and *rnf1* operons and the *fix* operon as a result of conditions favoring one of the three forms of nitrogenase compard to the control. (FC, Fold Change; Nif, Mo-dependent nitrogenase; Vnf, V-dependent nitrogenase; Anf, Fe-only nitrogenase)

Regulation					
Locus Tag	FC Nif/Control	FC Vnf/Control	FC Anf/Control	Gene name	Gene product
Avin_02780	0.1	2.9	2.7	*vnfA*	Vanadium nitrogenase σ^{54}-dependent transcriptional activator, VnfA
Avin_33440	4.6	23.7	24.3	*vnfA2*	σ^{54}-dependent activator protein
Avin_49020	0.4	2.2	4.6	*anfA*	σ^{54}-dependent transcriptional activator for the iron only nitrogenase, AnfA
Avin_51000	0.9	1.0	1.1	*nifA*	Nif-specific σ^{54}-dependent transcriptional activator protein, NifA
Avin_47100	0.6	5.8	5.2	*vnfA3*	σ^{54}-dependent activator protein

Electron Transport					
Locus Tag	FC Nif/Control	FC Vnf/Control	FC Anf/Control	Gene name	Gene product
Avin_10510	3.9	21.9	46.9	*fixP*	4Fe-4S ferredoxin, iron-sulfur binding domain protein
Avin_10520	8.5	65.2	34.0	*fixA*	Electron transfer flavoprotein beta-subunit, FixA
Avin_10530	40.5	231.0	206.2	*fixB*	Electron transfer flavoprotein, alpha subunit, FixB
Avin_10540	3.5	18.7	21.3	*fixC*	Electron-transferring-flavoprotein dehydrogenase, FixC
Avin_10550	16.2	46.8	52.0	*fixX*	FixX ferredoxin-like protein
Avin_19220	1.2	0.9	0.9	*rnfE*	Electron transport complex, subunit E
Avin_19230	1.5	0.8	1.4	*rnfG*	Electron transport complex, subunit G
Avin_19240	0.8	0.7	0.4	*rnfD*	Electron transport complex, subunit D
Avin_19250	3.1	1.5	1.9	*rnfC*	Electron transport complex, subunit C
Avin_19260	2.5	1.4	1.0	*rnfB*	Electron transport complex, subunit B
Avin_19270	1.0	1.0	0.6	*rnfA*	Electron transport complex, subunit A

Avin_50920	4.1	2.5	1.9	*rnfH*	Electron transport complex, RnfABCDGE type, H subunit
Avin_50930	11.6	7.8	6.7	*rnfE1*	Electron transport complex, RnfABCDGE type, E subunit
Avin_50940	8.5	7.7	8.6	*rnfG1*	Electron transport complex, RnfABCDGE type, G subunit
Avin_50950	8.7	5.5	4.9	*rnfD1*	Electron transport complex, RnfABCDGE type, D subunit
Avin_50960	15.5	10.5	10.7	*rnfC1*	Electron transport complex, RnfABCDGE type, C subunit
Avin_50970	13.3	7.8	8.1	*rnfB1*	Electron transport complex, RnfABCDGE type, B subunit
Avin_50980	39.4	14.2	12.0	*rnfA1*	Electron transport complex, RnfABCDGEFH, A subunit

Table D5. Changes in transcript levels of genes implicated in the global respones *of A. vinelandii* for Mo-independent nitrogen fixation. (FC, Fold Change; Vnf, V-dependent nitrogenase; Anf, Fe-only nitrogenase)

Molybdate Transport				
Locus Tag	FC Vnf/Control	FC Anf/Control	Gene name	Gene product
Avin_01280	2.4	10.3	-	ABC transporter, membrane spanning protein
Avin_01290	3.7	14.9	-	ABC transporter, membrane spanning protein
Avin_01300	2.7	4.1	*modA2*	Periplasmic molybdate-binding protein, ModA2
Avin_50650	3.8	6.0	*modC1*	Mo transporter, inner membrane ATP-binding component, ModC1
Avin_50660	6.2	5.4	*modB1*	Mo transporter membrane protein, ModB1
Avin_50670	8.3	7.1	*modA1*	Molybdenum transporter, periplasmic molybdate-binding protein
Avin_50680	4.9	5.3	*modE*	Mo regulation, Mo processing homeostasis
Avin_50690	17.3	17.5	*modG*	Mo processing, homeostasis
Avin_50700	7.4	8.7	-	ABC transporter ATP binding component
Avin_50710	1.4	2.8	-	ABC type tungstate transporter, permease component
Avin_50720	2.8	2.3	-	ABC transporter, permease protein
Avin_50730	4.8	6.4	*modA3*	Molybdate ABC transporter, periplasmic molybdate-binding protein
Pilus				
Locus Tag	FC Vnf/Control	FC Anf/Control	Gene name	Gene product
Avin_03090	2.4	11.0	*pilT*	Pilus retraction protein
Avin_03180	1.7	6.8	*pilG*	Type IV pili response regulator, PilG
Avin_11820	2.8	4.5	*fimT*	Type IV pilus biogenesis protein, FimT
Avin_11830	5.8	12.7	*pilE*	Type IV PilE-like pilus assembly protein
Avin_11840	2.8	23.2	*pilV*	Type IV pilus assembly transmembrane protein, PilV-like protein
Avin_11850	2.1	4.0	*pilW*	Type IV pilus assembly transmembrane protein, PilW-like protein
Avin_11870	2.0	3.8	*pilY1*	Type IV pilus assembly

				protein, PilY1-like protein
Avin_12070	2.7	3.2	*pilD*	Type IV Pilus Prepilin peptidase, PilD
Avin_12080	0.5	0.9	*pilC*	Type IV pilus biogenesis protein
Avin_12090	4.4	9.9	*pilB*	Type IV pilus assembly protein
Avin_12104	4.4	15.7	*pilA*	Fimbrial protein
Avin_45260	4.4	5.3	*pilQ*	Bacterial secretion system protein
Avin_45270	5.5	11.7	*pilP*	Pilus assembly protein
Avin_45280	1.7	3.2	*pilO*	Pilus assembly protein
Avin_45290	6.4	11.5	*pilN*	Fimbrial assembly protein
Avin_45300	8.3	20.3	*pilM*	Type IV pilus assembly protein
ISC Machinery				
Locus Tag	FC Vnf/Control	FC Anf/Control	Gene name	Gene product
Avin_40340	0.40	0.30	*iscX*	Conserved hypothetical protein
Avin_40350	0.36	0.27	*fdx*	Isc ferredoxin
Avin_40360	0.39	0.42	*hscA*	Fe-S protein assembly chaperone, HscA
Avin_40370	0.51	0.47	*hscB*	Co-chaperone Hsc20
Avin_40380	0.22	0.58	*iscA*	Iron-sulfur biogenesis protein
Avin_40390	0.18	0.19	*iscU*	Iron-sulfur cluster assembly scaffold protein
Avin_40400	0.22	0.21	*iscS*	Cysteine desulfurase IscS
Avin_40410	0.64	0.79	*iscR*	Iron-sulphur cluster assembly transcription factor, IscR
Other				
Avin_28760	0.20	0.15	*nfuA*	NfuA protein
Avin_46030	0.68	0.45	*erpA*	Essential respiratory protein A, ErpA
[NiFe]-Hydrogenase				
Locus Tag	FC Vnf/Control	FC Anf/Control	Gene name	Gene product
Avin_04360	11.1	8.3	-	Soluble hydrogenase, alpha or beta chain
Avin_04370	19.9	16.1	-	Cyclic nucleotide-binding protein, hydrogenase accessory protein, HoxI
Avin_04380	8.9	11.7	-	Soluble hydrogenase gamma subunit
Avin_04390	29.6	29.7	*hoxY*	Soluble hydrogenase delta subunit, HoxY
Avin_04400	8.3	11.8	*hoxH*	Soluble nickel-dependent hydrogenase, large subunit, HoxH

Avin_04410	5.8	18.2	*hoxW*	Soluble hydrogenase processing peptidase, HoxW
Avin_50440	4.2	3.7	*hypE*	Hydrogenase expression/formation protein, HypE
Avin_50450	1.3	1.4	*hypD*	Hydrogenase expression/formation protein, HypD
Avin_50460	3.6	2.3	*hypC*	Hydrogenase assembly chaperone, HypC/HupF
Avin_50470	4.5	3.4	*hypF*	Hydrogenase maturation carbamoyltransferase, HypF
Avin_50480	3.5	5.6	*hypB*	Hydrogenase nickel incorporation protein, HypB
Avin_50490	23.7	22.0	*hypA*	Hydrogenase nickel incorporation protein, HypA
Avin_50500	17.5	11.2	*hoxV*	Hydrogenase expression/formation protein, HoxV
Avin_50510	5.8	8.0	*hoxT*	Hydrogenase expression/formation protein, HoxT
Avin_50520	5.8	8.5	*hoxR*	Rubredoxin-type Fe(Cys)4 protein, HoxR
Avin_50530	6.7	5.8	*hoxQ*	Hydrogenase expression/formation protein, HoxQ
Avin_50540	2.9	3.1	*hoxO*	Hydrogenase expression/formation protein, HoxO
Avin_50550	2.2	3.2	*hoxL*	Hydrogenase assembly chaperone, HoxL (HypC/HupF family)
Avin_50560	2.2	5.1	*hoxM*	Hydrogenase expression/formation protein, HoxM
Avin_50570	8.1	5.2	*hoxZ*	Ni/Fe-hydrogenase, b-type cytochrome subunit, HoxZ
Avin_50580	4.6	3.4	*hoxG*	Membrane bound nickel-dependent hydrogenase, large subunit, HoxG
Avin_50590	5.3	3.0	*hoxK*	Uptake hydrogenase small subunit (Precursor), HoxK
Terminal Oxidases				
Locus Tag	FC Vnf/Control	FC Anf/Control	Gene name	Gene product
Avin_00950	20.3	23.2	-	Cytochrome oxidase assembly protein

Avin_00960	4.5	5.9	-	Conserved hypothetical protein
Avin_00970	7.1	8.9	-	Surfeit locus 1-like protein
Avin_00980	7.9	6.0	-	Conserved hypothetical protein
Avin_00990	4.4	5.9	-	Cytochrome C Oxidase, Subunit III
Avin_01000	8.0	12.9	-	Cytochrome c oxidase assembly protein CtaG/Cox11
Avin_01010	8.8	9.7	*ctaD*	Cytochrome c oxidase, subunit I
Avin_01020	17.7	18.4	-	Cytochrome c oxidase, subunit II
Avin_11040	0.3	0.3	-	Cytochrome bd ubiquinol oxidase, subunit II
Avin_11050	1.4	2.3	-	Cytochrome bd ubiquinol oxidase, subunit I
Avin_11170	1.7	28	*coxB*	Cytochrome o ubiquinol oxidase, subunit II
Avin_11180	7.8	7.7	*coxA*	Cytochrome o ubiquinol oxidase, subunit I
Avin_13060	1.0	0.8	*petA*	Ubiquinol-cytochrome c reductase, iron-sulfur subunit
Avin_13070	0.9	1.0	*petB*	Cytochrome b/b6
Avin_13080	2.2	2.0	*petC*	Ubiquinol--cytochrome c reductase, cytochrome c1
Avin_19880	4.3	4.7	*cydB*	Cytochrome bd ubiquinol oxidase, subunit II
Avin_19890	3.7	4.8	*cydA*	Cytochrome bd ubiquinol oxidase, subunit I
Avin_19910	1.2	1.2	*cydR*	Fnr-like negative transcriptional regulator of CydAB
Avin_19940	1.6	1.3	*ccoS*	cbb3-type cytochrome oxidase maturation protein
Avin_19950	2.1	1.9	*ccoI*	Copper-translocating P-type ATPase
Avin_19960	2.5	2.8	*ccoH*	Cytochrome c oxidase accessory protein, CcoH
Avin_19970	1.8	1.4	*ccoG*	Cytochrome c oxidase accessory protein, CcoG
Avin_19980	1.8	1.3	*ccoP*	Cytochrome c oxidase, cbb3-type, subunit III

Avin_19990	2.3	1.7	*ccoQ*	Cytochrome c oxidase, cbb3-type, subunit IV
Avin_20000	1.8	1.5	*ccoO*	Cytochrome c oxidase, cbb3-type, subunit II
Avin_20010	1.1	1.0	*ccoN*	Cytochrome c oxidase, cbb3-type, subunit I

NADH-Ubiquinone Oxidoreductases

Locus Tag	FC Vnf/Control	FC Anf/Control	Gene name	Gene product
Avin_12000	2.5	2.8	*ndh*	Uncoupled NADH:quinone oxidoreductase
Avin_14590	1.4	1.2	*nqrA*	Na-translocating NADH-ubiquinone oxidoreductase protein, subunit A
Avin_14600	1.1	1.2	*nqrB*	NADH:ubiquinone oxidoreductase, Na(+)-translocating protein, subunit B
Avin_14610	1.7	1.4	*nqrC*	NADH:ubiquinone oxidoreductase, Na(+)-translocating protein, subunit C
Avin_14620	1.2	1.0	*nqrD*	NADH:ubiquinone oxidoreductase, Na(+)-translocating protein, subunit D
Avin_14630	1.5	1.3	*nqrE*	NADH:ubiquinone oxidoreductase, Na(+)-translocating protein subunit E
Avin_14640	1.4	1.3	*nqrF*	NADH:ubiquinone oxidoreductase, Na(+)-translocating protein, subunit F
Avin_19530	6.7	4.5	*shaAB*	Sodium hydrogen antiporter subunitAB, ShaAB
Avin_19540	3.8	5.7	*shaC*	Sodium hydrogen antiporter subunitC, ShaC
Avin_19550	9.3	7.7	*shaD*	Sodium hydrogen antiporter subunitD, ShaD
Avin_19560	1.8	1.8	*shaE*	Sodium hydrogen antiporter subunitE, ShaE
Avin_19570	2.8	2.3	*shaF*	Sodium hydrogen antiporter subunitF, ShaF
Avin_19580	8.9	6.7	*shaG*	Sodium hydrogen antiporter subunitG, ShaG
Avin_28440	2.1	2.2	*nuoA*	NADH-ubiquinone oxidoreductase, chain

				alpha
Avin_28450	4.2	3.7	*nuoB*	NADH-ubiquinone oxidoreductase, chain beta
Avin_28460	2.2	2.1	*nuocd*	NADH-ubiquinone oxidoreductase, chain cd
Avin_28470	1.7	1.5	*nuoE*	NADH-ubiquinone oxidoreductase, chain E
Avin_28480	1.7	1.3	*nuoF*	NADH-quinone oxidoreductase, F subunit
Avin_28490	2.1	1.6	*nuoG*	NADH-quinone oxidoreductase, chain G
Avin_28500	1.5	1.6	*nuoH*	Respiratory-chain NADH dehydrogenase, subunit 1, NuoH
Avin_28510	1.7	2.2	*nuoI*	NADH-quinone oxidoreductase, chain I
Avin_28520	1.2	1.3	*nuoJ*	NADH-ubiquinone/plastoquinone oxidoreductase, chain J
Avin_28530	2.3	1.9	*nuoK*	NADH-ubiquinone oxidoreductase, chain K
Avin_28540	1.4	1.2	*nuoL*	NADH-plastoquinone oxidoreductase, chain L
Avin_28550	1.4	1.7	*nuoM*	Proton-translocating NADH-quinone oxidoreductase, chain M
Avin_28560	1.6	1.6	*nuoN*	Proton-translocating NADH-quinone oxidoreductase, chain N

ATP Synthases

Locus Tag	FC Vnf/Control	FC Anf/Control	Gene name	Gene product
Avin_19670	2.5	4.1	-	ATP synthase F1, beta subunit
Avin_19680	2.1	3.5	-	H+-transporting two-sector ATPase, delta/epsilon subunit
Avin_19690	2.1	2.0	-	H(+)-transporting ATP synthase, gene 1
Avin_19700	1.0	1.8	-	lipoprotein
Avin_19710	1.5	1.6	-	H+-transporting two-sector ATPase, A subunit
Avin_19720	1.4	1.2	-	ATP synthase F0, C subunit
Avin_19730	1.9	1.8	-	H+-transporting two-sector ATPase, B/B subunit
Avin_19740	2.8	3.8	-	ATP synthase F1, alpha subunit
Avin_19750	1.4	4.2	-	H+-transporting two-

				sector ATPase, gamma subunit
Avin_52150	1.1	1.2	*atpC*	F1 sector of membrane-bound ATP synthase, epsilon subunit
Avin_52160	1.6	2.0	*atpD*	ATP synthase F1, beta subunit
Avin_52170	1.2	1.1	*atpG*	F1 sector of membrane-bound ATP synthase, gamma subunit
Avin_52180	1.7	2.2	*atpA*	ATP synthase F1, alpha subunit
Avin_52190	1.1	1.1	*atpH*	F1 sector of membrane-bound ATP synthase, delta subunit
Avin_52200	1.3	1.5	*atpF*	F0 sector of membrane-bound ATP synthase, subunit B
Avin_52210	2.1	2.4	*atpE*	F0 sector of membrane-bound ATP synthase, subunit C
Avin_52220	1.5	2.2	*atpB*	F0 sector of membrane-bound ATP synthase, subunit A
Avin_52230	3.1	4.2	*atpI*	ATP synthase I chain

Table D6. Fold change in transcript abundance of the top ten up- and down-regulated genes identified in pairwise comparison of experimental conditions. (FC, Fold Change; Nif, Mo-dependent nitrogenase; Vnf, V-dependent nitrogenase; Anf, Fe-only nitrogenase)

Increase			
Nif/Control			
Locus Tag	FC	Gene name	Gene product
Avin_45300	214.2	*pilM*	Type IV pilus assembly protein
Avin_11830	138.5	*pilE*	Type IV PilE-like pilus assembly protein
Avin_11860	110.3	-	Conserved hypothetical protein
Avin_03180	105.2	*pilG*	Type IV pili response regulator, PilG
Avin_12104	97.9	*pilA*	Fimbrial protein
Avin_11840	94.4	*pilV*	Type IV pilus assembly transmembrane protein, PilV-like protein
Avin_02940	93.0	*coaBC*	P-pantothenate cysteine ligase / P-pantothenoylcysteine decarboxylase
Avin_45290	90.7	*pilN*	Fimbrial assembly protein
Avin_43500	82.4	-	Cyclic 3,5-nucleotide monophosphate metallophosphodiesterase
Avin_45270	66.6	*pilP*	Pilus assembly protein
Avin_20560	65.9	-	Hypothetical protein
Avin_03090	60.3	*pilT*	Pilus retraction protein
Avin_18070	53.9	-	ATPase, AAA superfamily
Avin_00180	44.7	*lysM*	Peptidoglycan-binding LysM protein
Avin_34830	43.3	-	Hypothetical protein
Increase			
Vnf/Control			
Locus Tag	FC	Gene name	Gene product
Avin_10530	231.0	*fixB*	Electron transfer flavoprotein, alpha subunit, FixB
Avin_16830	137.2	-	Hypothetical protein
Avin_02550	101.8	-	Phosphonate transport system substrate-binding protein
Avin_02410	94.9	-	ABC transporter, ATP-binding component
Avin_47120	87.7	-	Conserved hypothetical protein
Avin_46150	76.7	-	Conserved hypothetical protein
Avin_10520	65.2	*fixA*	Electron transfer flavoprotein beta-subunit, FixA
Avin_02560	65.0	-	ABC transporter
Avin_02530	57.4	-	Hypothetical protein
Avin_02540	53.2	-	Phosphonate transport system inner membrane component
Avin_02430	50.4	-	Hypothetical protein
Avin_10550	46.8	*fixX*	FixX ferredoxin-like protein
Avin_46160	41.6	*pdhB*	Pyruvate dehydrogenase E1 beta subunit
Avin_26180	39.9	*hutU*	Urocanate hydratase
Avin_26170	39.9	-	Hypothetical protein
Increase			
Anf/Control			
Locus Tag	FC	Gene name	Gene product
Avin_10530	206.2	*fixB*	Electron transfer flavoprotein, alpha subunit, FixB
Avin_05810	95.1	-	Hypothetical protein
Avin_46150	91.9	-	Conserved hypothetical protein

Avin_47120	90.1	-	Conserved hypothetical protein
Avin_16830	82.9	-	Hypothetical protein
Avin_05820	59.5	-	Hypothetical protein
Avin_10550	52.0	*fixX*	FixX ferredoxin-like protein
Avin_46160	48.4	*pdhB*	Pyruvate dehydrogenase E1 beta subunit
Avin_10510	46.0	-	4Fe-4S ferredoxin, iron-sulfur binding domain protein
Avin_29510	43.9	-	ABC-type antimicrobial peptide transport system, ATPase component
Avin_18070	41.4	-	ATPase, AAA superfamily
Avin_17910	40.4	-	Glycosyl transferase, group 1 family protein
Avin_26170	40.3	-	Hypothetical protein
Avin_26180	38.1	*hutU*	Urocanate hydratase
Avin_00800	35.9	-	Conserved hypothetical protein

Decrease

Nif/Control

Locus Tag	FC	Gene name	Gene product
Avin_01050	0.0	-	Sulfate transporter
Avin_47210	0.0	-	Alpha/beta fold hydrolase
Avin_15290	0.0	*ohr*	Organic hydroperoxide resistance protein (OsmC family)
Avin_09340	0.0	-	Conserved hypothetical protein
Avin_48600	0.0	*pstS*	Phosphate-binding protein PstS
Avin_47230	0.0	-	Conserved hypothetical protein
Avin_09330	0.0	-	Siderophore biosynthesis protein, IucA/IucC family
Avin_26370	0.0	-	Hypothetical protein
Avin_26350	0.0	-	Aldo/keto reductase protein
Avin_21180	0.0	*entA*	2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase, short-chain dehydrogenase/reductase
Avin_01060	0.0	-	Carbonic anhydrase
Avin_19060	0.0	-	Hypothetical protein
Avin_22850	0.0	-	Conserved hypothetical protein 156
Avin_01340	0.0	-	AAS bifunctional protein
Avin_49410	0.0	-	Conserved hypothetical protein

Decrease

Vnf/Control

Locus Tag	FC	Gene name	Gene product
Avin_22850	0.0	-	Conserved hypothetical protein 156
Avin_09340	0.0	-	Conserved hypothetical protein
Avin_49410	0.0	-	Conserved hypothetical protein
Avin_01050	0.0	-	Sulfate transporter
Avin_47230	0.0	-	Conserved hypothetical protein
Avin_09330	0.0	-	Siderophore biosynthesis protein, IucA/IucC family
Avin_47210	0.0	-	Alpha/beta fold hydrolase
Avin_01060	0.0	-	Carbonic anhydrase
Avin_09320	0.0	-	Transporter, MFS superfamily
Avin_40120	0.0	-	2-isopropylmalate synthase
Avin_22840	0.0	-	SdiA-regulated domain-containing protein
Avin_21230	0.0	*csbX*	Catecholate siderophore efflux pump, MFS_1 family
Avin_09300	0.0	-	Diaminopimelate decarboxylase
Avin_09310	0.0	-	Siderophore biosynthesis protein, IucA/IucC family

Avin_48600	0.0	*pstS*	Phosphate-binding protein PstS
Decrease			
Anf/Control			
Locus Tag	FC	Gene name	Gene product
Avin_49410	0.0	-	Conserved hypothetical protein
Avin_22850	0.0	-	Conserved hypothetical protein 156
Avin_09330	0.0	-	Siderophore biosynthesis protein, IucA/IucC family
Avin_09340	0.0	-	Conserved hypothetical protein
Avin_09310	0.0	-	Siderophore biosynthesis protein, IucA/IucC family
Avin_01050	0.0	-	Sulfate transporter
Avin_47230	0.0	-	Conserved hypothetical protein
Avin_47210	0.0	-	Alpha/beta fold hydrolase
Avin_01060	0.0	-	Carbonic anhydrase
Avin_09320	0.0	-	Transporter, MFS superfamily
Avin_48600	0.0	*pstS*	Phosphate-binding protein PstS
Avin_22840	0.0	-	SdiA-regulated domain-containing protein
Avin_21230	0.0	*csbX*	Catecholate siderophore efflux pump, MFS_1 family
Avin_40120	0.0	-	2-isopropylmalate synthase
Avin_36980	0.0	-	Hypothetical protein

Table D7. SOLiD™ sequencing information.

Treatment	Mapped Reads	% Mapped Reads	Mapped in rDNA Regions	% rDNA	Remaining Mapped Reads	Uniquely Mapped Reads	% Unique Reads
Control	22,783,073.00	39.9%	22342963	98.1%	880,220	844,536	97.9%
Nif	24,432,540.00	55.3%	23343369	95.5%	1,089,171	1,029,518	93.9%
Vnf	28,259,018.00	49.8%	26534353	93.9%	1,724,665	1,633,893	94.4%
Anf	32,349,986.00	52.7%	30048661	92.9%	2,301,325	2,174,981	94.2%

Table D8. Fifteen most abundant mRNA species in each condition. (Nif, Mo-dependent nitrogenase; Vnf, V-dependent nitrogenase; Anf, Fe-only nitrogenase)

Control			
Locus tag	Counts	Gene name	Gene product
Avin_47230	29335	-	Conserved hypothetical protein
Avin_16040	27280	-	Hypothetical protein
Avin_44900	17548	*aceE*	Pyruvate dehydrogenase, E1 component
Avin_48330	11709	*serA*	D-3-phosphoglycerate dehydrogenase
Avin_13480	10888	*groEL*	Chaperonin GroEL/Hsp60
Avin_28310	10728	*icd*	Isocitrate dehydrogenase, NADP-dependent, Icd
Avin_06220	8197	*fusA*	Translation elongation factor G
Avin_15870	7955	*rpsA*	30S ribosomal protein S1
Avin_13980	7482	*metE*	5-methyltetrahydropteroyltriglutamate--homocysteine S-methyltransferase
Avin_29770	6808	*sucA*	2-oxoglutarate dehydrogenase, E1 component
Avin_23590	6806	*pepS16*	Peptidase S16, ATP-dependent protease
Avin_23470	6698	*acnB*	Aconitate hydratase 2
Avin_23330	6688	*oprF*	Major outer membrane porin OprF
Avin_19890	6076	*cydA*	Cytochrome bd ubiquinol oxidase, subunit I
Avin_22850	5940	-	Conserved hypothetical protein 156
Nif			
Locus tag	Counts	Gene name	Gene product
Avin_01400	23928	*nifK*	Nitrogenase MoFe protein beta chain, NifK
Avin_01390	20133	*nifD*	Nitrogenase MoFe protein alpha chain, NifD
Avin_01380	18280	*nifH*	Nitrogenase Fe protein
Avin_19890	13884	*cydA*	Cytochrome bd ubiquinol oxidase, subunit I
Avin_01360	12458	-	Conserved hypothetical protein
Avin_38680	10080	*rpoS*	RNA polymerase sigma factor
Avin_19880	9268	*cydB*	Cytochrome bd ubiquinol oxidase, subunit II
Avin_19860	6020	-	Hypothetical protein
Avin_48330	5866	*serA*	D-3-phosphoglycerate dehydrogenase
Avin_16040	5810	-	Hypothetical protein
Avin_20450	4920	*rplT*	50S ribosomal protein L20
Avin_15870	4833	*rpsA*	30S ribosomal protein S1
Avin_20440	4370	*rpmI*	50S ribosomal protein L35
Avin_06440	4328	*rplO*	50S ribosomal protein L15
Avin_06220	4190	*fusA*	Translation elongation factor G
Vnf			
Locus tag	Counts	Gene name	Gene product
Avin_02660	47108	*vnfH*	Vanadium nitrogenase Fe protein
Avin_23670	23207	*phbP*	Phasin protein
Avin_02610	21269	*vnfD*	Vanadium nitrogenase, alpha subunit, VnfD
Avin_23640	21009	*phbA*	Polyhydroxybutyrate biosynthetic beta-ketothiolase
Avin_02590	19560	*vnfK*	Vanadium nitrogenase beta subunit, VnfK
Avin_23650	16255	*phbB*	Acetoacetl-CoA reductase in PHB biosynthesis
Avin_38680	10086	*rpoS*	RNA polymerase sigma factor
Avin_19890	9524	*cydA*	Cytochrome bd ubiquinol oxidase, subunit I
Avin_19880	8469	*cydB*	Cytochrome bd ubiquinol oxidase, subunit II
Avin_44900	8105	*aceE*	Pyruvate dehydrogenase, E1 component
Avin_16040	8060	-	Hypothetical protein
Avin_23330	7416	*oprF*	Major outer membrane porin OprF

Avin_37820	6074	*sodB*	Fe-Superoxide dismutase
Avin_19330	5879	-	Hypothetical protein
Avin_25770	5307	-	Heat shock Hsp20 protein

Anf

Locus tag	Counts	Gene name	Gene product
Avin_49000	47166	*anfH*	Nitrogenase Fe protein, AnfH
Avin_23670	20619	*phbP*	Phasin protein
Avin_16040	16955	-	Hypothetical protein
Avin_23640	16854	*phbA*	Polyhydroxybutyrate biosynthetic beta-ketothiolase
Avin_48990	15893	*anfD*	Fe-only nitrogenase, alpha subunit, NifD
Avin_23650	13006	*phbB*	Acetoacetl-CoA reductase in PHB biosynthesis
Avin_48970	12930	*anfK*	Fe-only nitrogenase, beta subunit, NifK
Avin_19890	10066	*cydA*	Cytochrome bd ubiquinol oxidase, subunit I
Avin_19880	7798	*cydB*	Cytochrome bd ubiquinol oxidase, subunit II
Avin_44900	6823	*aceE*	Pyruvate dehydrogenase, E1 component
Avin_25770	6716	-	Heat shock Hsp20 protein
Avin_38680	6666	*rpoS*	RNA polymerase sigma factor
Avin_37820	5525	*sodB*	Fe-Superoxide dismutase
Avin_23330	5485	*oprF*	Major outer membrane porin OprF
Avin_19330	5264	-	Hypothetical protein
Avin_23630	5101	*phbC*	PHB synthase
Avin_13300	5064	*ftsZ*	Cell division protein FtsZ

Table D9. Trancsript counts of differentially expressed genes chosen for qRT-PCR analysis to verify SOLiD™ sequencing results. (Nif, Mo-dependent nitrogenase; Vnf, V-dependent nitrogenase; Anf, Fe-only nitrogenase)

Locus Tag	Gene name	Control	Nif	Vnf	Anf	Gene products
Avin_01380	*nifH*	127	18280	322	331	Nitrogenase Fe protein, NifH
Avin_01400	*nifK*	434	23928	949	786	Nitrogenase MoFe protein beta chain, NifK
Avin_02600	*vnfG*	218	130	4790	206	Vanadium nitrogenase, delta subunit, VnfG
Avin_02770	*vnfE*	194	188	5368	5190	Nitrogenase VFe cofactor biosynthesis protein, VnfE
Avin_08270	*ftsH*	218	130	143	168	ATP-dependent metallopeptidase M41, FtsH
Avin_09310	-	204	7	3	1	Siderophore biosynthesis protein, IucA/IucC family
Avin_10530	*fixB*	2	105	554	495	Electron transfer flavoprotein, alpha subunit, FixB
Avin_19240	*rnfD*	35	78	69	51	Electron transport complex, subunit D, RnfD
Avin_26180	*hutU*	7	27	287	224	Urocanate hydratase
Avin_34460	*alaS*	338	371	233	309	Alanyl-tRNA synthetase
Avin_40400	*iscS*	895	96	194	190	Cysteine desulfurase, IscS
Avin_45300	*pilM*	5	1028	48	163	Type IV pilus assembly protein, PilM
Avin_47000	*rpoD*	1226	831	1149	1165	RNA polymerase sigma factor, sigma70; RpoD
Avin_47230	Hyp	29335	200	187	208	Conserved hypotheitcal protein
Avin_48600	*pstS*	4364	26	76	66	Phosphate-binding protein, PstS (PhoT family)
Avin_48980	*anfG*	278	18	14	1906	Fe-only nitrogenase, delta subunit, AnfG
Avin_50580	*hoxG*	74	333	345	252	Membrane bound nickel-dependent hydrogenase, large subunit, HoxG
Avin_50980	*rnfA1*	31	1227	444	374	Electron transport complex, RnfABCDGEFH, A subunit, RnfA

Avin_51010	*nifB*	74	2324	1542	1584	Nitrogenase cofactor biosynthesis protein, NifB
Avin_51250	algE7	91	2686	772	894	Mannuronan C-5 epimerase/alginate lyase

Table D10. Genes targeted and primers used for RT-PCR and qRT-PCR analysis.

Locus Tag	Gene name	Gene products	Forward Primer (5' - 3')	Reverse Primer (5' - 3')
Avin_01380	*nifH*	Nitrogenase Fe protein, NifH	CATCTACGGCA AAGGTGGTATC GGTAAGTC	GATGTAGATTT CCTGGGCCTTG TTCTCG
Avin_01400	*nifK*	Nitrogenase MoFe protein beta chain, NifK	CGGCTTCGAGG AAAAGTATCC GCAG	GTAGCATTCTT GCCGTAGGGCG AAG
Avin_02600	*vnfG*	Vanadium nitrogenase, delta subunit, VnfG	ATGAGCCAGTC CCATCTCGACG ATC	TCAGTAGAGGT GGTGGTTGAGT TCGC
Avin_02770	*vnfE*	Nitrogenase VFe cofactor biosynthesis protein, VnfE	CAGCTTCTACG GTATCGCCGAT ACCTC	CTCTGGTCGTT GTCGTCGATCA TCC
Avin_08270	*ftsH*	ATP-dependent metallopeptidase M41, FtsH	GAAACAACAG GCGCAGTTCCA TATCGGTTAC	CAGAGGCTGAA GAAGATCAGCG CC
Avin_09310	-	Siderophore biosynthesis protein, IucA/IucC family	CATTGAGCCAT TTCCTGAAGTT CTCGATCC	GAGGTTGTTGA CCAGGCTGCAG TAG
Avin_10530	*fixB*	Electron transfer flavoprotein, alpha subunit, FixB	CTTCGTCGATT ACCGTCACGTC TG	GATTTCGTGCT GCAGCAGATAA GCC
Avin_19240	*rnfD*	Electron transport complex, subunit D, RnfD	GACTTGGCTGT ATGGTTTCGCT CCC	GTCCATCGTCA GGCTTCGATTG GTCTTC
Avin_26180	*hutU*	Urocanate hydratase	CTGTTGATGAA CAACCTCGATC CCGAG	CTCGACGAAGG TCTCGAAGGTG C
Avin_34460	*alaS*	Alanyl-tRNA synthetase	CGGCAAGCAC AACGATCTGG AAAATGTC	CTGCATGAACA CGTTGTTCCAG ATCTCGATGTA G
Avin_40400	*iscS*	Cysteine desulfurase, IscS	GATTTATCTGG ATTATTCCGCC ACCACTCC	GTGATGATGTG CTTGCCCTTGCT C
Avin_45300	*pilM*	Type IV pilus assembly protein, PilM	GATATCGGTTC GACCATGGTCA AACTTCTC	CATCTCCATGA CCTTGGTGATC ACCAC
Avin_47000	*rpoD*	RNA polymerase sigma factor, sigma70; RpoD	GTAAAAGCTC AACAGCAGTCT CGCATCAAAG	CTTGGCGATCT CGATTTCGCCTT C
Avin_47230	Hyp	Conserved hypotheitcal protein	GACCATCCAGC CGTTCGAGACG CTCAAC	CGCAGTTCGAC CTGGTAGATGT AGC
Avin_48600	*pstS*	Phosphate-binding protein, PstS (PhoT family)	CTACTTCAAGG AAGAGGCCCT CTGTAAAG	GCAGTGGAATA TAGCCGTCCTT GAC
Avin_48980	*anfG*	Fe-only nitrogenase, delta subunit, AnfG	CTGCTGTGGTC AAACAGAAGG TCGAAG	GTGTTTGTCGCT CAACTCTTCGTT GAGC
Avin_50580	*hoxG*	Membrane bound nickel-dependent hydrogenase, large subunit, HoxG	CCTGATCCGCA ACCTGATGGAC AAG	GACGATGTCCT TCTGCAGGTCC AG

Avin_50980	*rnfA1*	Electron transport complex, RnfABCDGEFH, A subunit, RnfA	GTTCTGTTTCT ATTGGGCACCG TGTTG	GATGTAGGCGA GGATGCGGATG AAG
Avin_51010	*nifB*	Nitrogenase cofactor biosynthesis protein, NifB	GATAGCGGCTT CACTTTTGAAG TCTGC	GCTGCCCGTTG ATCGACGAAAT ATTGATAAC
Avin_51250	*algE7*	Mannuronan C-5 epimerase/alginate lyase	GAAGGCGTGC TGATCAAGATG AGCAC	GTCCTTCTCGTA GCTGCGCAGAT AG

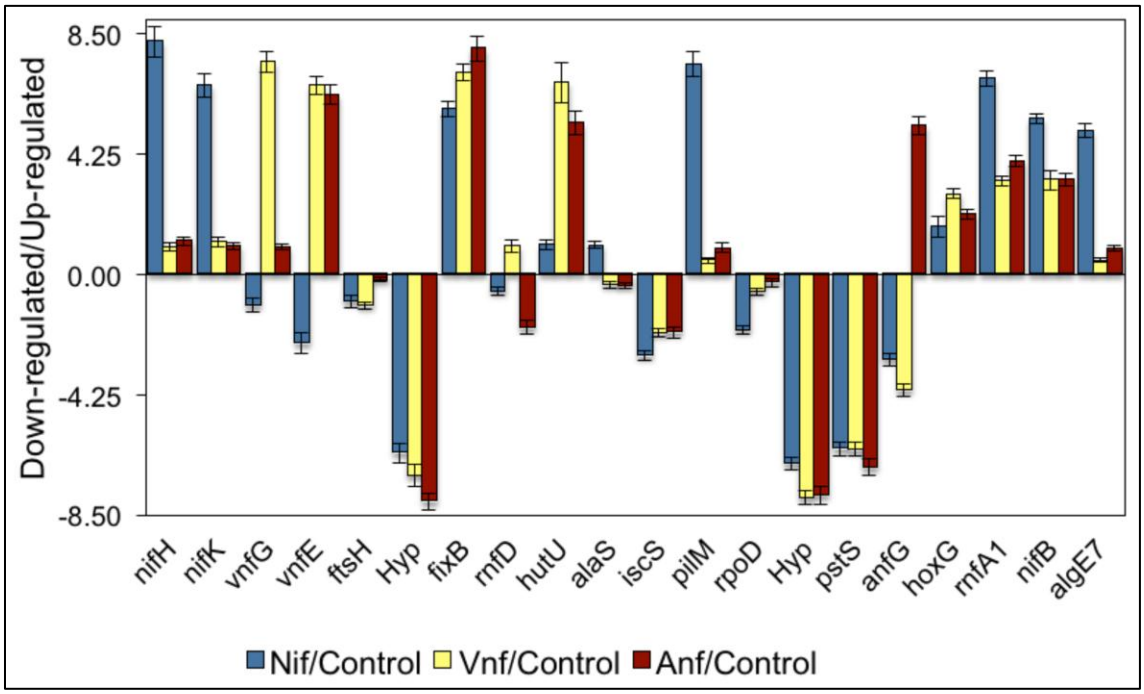


Figure D1. Results of qRT-PCR analyses. Results are presented as log(fold change). Targets are listed in Table D10. Assays were run in triplicate. (Nif, Mo-dependent nitrogenase; Vnf, V-dependent nitrogenase; Anf, Fe-only nitrogenase)

LITERATURE CITED

Abascal, F., R. Zardoya, et al. (2005). ProtTest: selection of best-fit models of protein evolution. Bioinformatics. 2 (9): 2104-2105.

Affourtit, J., J. P. Zehr, et al. (2001). Distribution of Nitrogen-Fixing Microorganisms along the Neuse River Estuary, North Carolina. Microb Ecol. 41 (2): 114-123.

Allen, R. M., M. J. Homer, et al. (1993). Dinitrogenase reductase- and MgATP-dependent maturation of apodinitrogenase from Azotobacter vinelandii. J Biol Chem. 268 (31): 23670-23674.

Amend, J. P. and E. L. Shock (2001). Energetics of overall metabolic reactions of thermophilic and hyperthermophilic Archaea and bacteria. FEMS Microbiol Rev. 25 (2): 175-243.

Anbar, A. D. (2008). Elements and evolution. Science. 322: 1481-1483.

Anbar, A. D., Y. Duan, et al. (2007). A whiff of oxygen before the great oxidation event? Science. 317 (5846): 1903-1906.

Anbar, A. D. and A. H. Knoll (2002). Proterozoic ocean chemistry and evolution: a bioinorganic bridge? Science. 297 (5584): 1137-1142.

Anders, S. and W. Huber Differential expression analysis for sequence count data. Genome Biol. 11 (10): R106.

Bageshwar, U. K., R. Raina, et al. (1998). Analysis of upstream activation of the vnfH promoter of Azotobacter vinelandii. Can J Microbiol. 44 (5): 405-415.

Barns, S. M., C. F. Delwiche, et al. (1996). Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. Proc Natl Acad Sci U S A. 93 (17): 9188-9193.

Beaumont, V. and F. Robert (1999). Nitrogen isotope ratios of keragens in Precambrian cherts: a record of the evolution of atmospheric chemistry? Precambrian Res. 96: 63-82.

Beiko, R. G., T. J. Harlow, et al. (2005). Highways of gene sharing in prokaryotes. Proc. Natl. Acad. Sci. USA. 102 (40): 14332-14337.

Bekker, A., H. D. Holland, et al. (2004). Dating the rise of atmospheric oxygen. Nature. 427 (6970): 117-120.

Belay, N., R. Sparling, et al. (1984). Dinitrogen fixation by a thermophilic methanogenic bacterium. Nature. 312 (5991): 286-288.

Bennett, L. T., M. R. Jacobson, et al. (1988). Isolation, sequencing, and mutagenesis of the nifF gene encoding flavodoxin from Azotobacter vinelandii. J Biol Chem. 263 (3): 1364-1369.

Bertsova, Y. V., A. V. Bogachev, et al. (2001). Noncoupled NADH:ubiquinone oxidoreductase of Azotobacter vinelandii is required for diazotrophic growth at high oxygen concentrations. J Bacteriol. 183 (23): 6869-6874.

Bishop, P. E., M. E. Hawkins, et al. (1986). Nitrogen fixation in molybdenum-deficient continuous culture by a strain of Azotobacter vinelandii carrying a deletion of the structural genes for nitrogenase (nifHDK). Biochem J. 238 (2): 437-442.

Bishop, P. E., D. M. Jarlenski, et al. (1980). Evidence for an alternative nitrogen fixation system in Azotobacter vinelandii. Proc Natl Acad Sci U S A. 77 (12): 7342-7346.

Bishop, P. E., D. M. Jarlenski, et al. (1982). Expression of an alternative nitrogen fixation system in Azotobacter vinelandii. J Bacteriol. 150 (3): 1244-1251.

Bolin, J. T., A. E. Ronco, et al. (1993). The unusual metal clusters of nitrogenase: structural features revealed by x-ray anomalous diffraction studies of the MoFe protein from Clostridium pasteurianum. Proc Natl Acad Sci U S A. 90 (3): 1078-1082.

Boone, D. R., W. B. Whitman, et al., Eds. (1993). Methanogenesis. New York & London, Chapman & Hall.

Boyd, E. S., A. D. Anbar, et al. (2009). An Early Origin for Molybdenum-Nitrogenase. Science. In review.

Boyd, E. S., A. D. Anbar, et al. (2011). A late methanogen origin for molybdenum-dependent nitrogenase. Geobiol. 9: 221–232.

Boyd, E. S., D. E. Cummings, et al. (2007). Mineralogy influences structure and diversity of bacterial communities associated with geological substrata in a pristine aquifer. Microb Ecol. 54 (1): 170-182.

Boyd, E. S., T. L. Hamilton, et al. (2010). [FeFe]-hydrogenase in Yellowstone National Park: evidence for dispersal limitation and phylogenetic niche conservatism. ISME J. Epub 2010 Jun11 doi:10.1038/ismej.2010.76.

Boyd, E. S., R. A. Jackson, et al. (2007). Isolation, characterization, and ecology of sulfur-respiring crenarchaea inhabiting acid-sulfate-chloride-containing geothermal springs in Yellowstone National Park. Appl Environ Microbiol. 73 (20): 6669-6677.

Boyd, E. S., W. D. Leavitt, et al. (2009). CO(2) uptake and fixation by a thermoacidophilic microbial community attached to precipitated sulfur in a geothermal spring. Appl Environ Microbiol. 75 (13): 4289-4296.

Boyd, E. S., J. R. Spear, et al. (2009). [FeFe] hydrogenase genetic diversity provides insight into molecular adaptation in a saline microbial mat community. Appl Environ Microbiol. 75 (13): 4620-4623.

Brigle, K. E., W. E. Newton, et al. (1985). Complete nucleotide sequence of the Azotobacter vinelandii nitrogenase structural gene cluster. Gene. 37 (1-3): 37-44.

Brigle, K. E., R. A. Setterquist, et al. (1987). Site-directed mutagenesis of the nitrogenase MoFe protein of Azotobacter vinelandii. Proc Natl Acad Sci U S A. 84 (20): 7066-7069.

Brill, W. J. (1980). Biochemical genetics of nitrogen fixation. Microbiol Rev. 44 (3): 449-467.

Brocker, M. J., S. Schomburg, et al. (2010). Crystal structure of the nitrogenase-like dark operative protochlorophyllide oxidoreductase catalytic complex (ChlN/ChlB)2. J Biol Chem. 285 (35): 27336-27345.

Brocker, M. J., S. Virus, et al. (2008). ATP-driven reduction by dark-operative protochlorophyllide oxidoreductase from Chlorobium tepidum mechanistically resembles nitrogenase catalysis. J Biol Chem. 283 (16): 10559-10567.

Brocker, M. J., D. Watzlich, et al. (2008). Substrate recognition of nitrogenase-like dark operative protochlorophyllide oxidoreductase from Prochlorococcus marinus. J Biol Chem. 283 (44): 29873-29881.

Brocks, J. J., R. Buick, et al. (2003). Composition and syngeneity of molecular fossils from the 2.78 to 2.45 billion-year-old Mount Bruce Supergroup, Pilbara Craton, Western Australia. Geochim. Cosmochim. Acta. 67: 4289–4319.

Bryant, J. A., C. Lamanna, et al. (2008). Colloquium paper: microbes on mountainsides: contrasting elevational patterns of bacterial and plant diversity. Proc Natl Acad Sci U S A. 105 Suppl 1: 11505-11511.

Buchanan-Wollaston, V., M. C. Cannon, et al. (1981). Role of the nifA gene product in the regulation of nif expression in Klebsiella pneumoniae. Nature. 294 (5843): 776-778.

Bulen, W. A. and J. R. LeComte (1966). The nitrogenase system from *Azotobacter*: two enzyme requirements for N_2 reduction, ATP dependent H_2 evolution and ATP hydrolysis. Proc. Natl. Acad. Sci. USA. 56: 979-986.

Burgess, B. K. and D. J. Lowe (1996). Mechanism of Molybdenum Nitrogenase. Chem Rev. 96 (7): 2983-3012.

Burke, D. H., J. E. Hearst, et al. (1993). Early evolution of photosynthesis: clues from nitrogenase and chlorophyll iron proteins. Proc Natl Acad Sci U S A. 90 (15): 7134-7138.

Burns, J. A., J. P. Zehr, et al. (2002). Nitrogen-fixing phylotypes of Chesapeake Bay and Neuse River estuary sediments. Microb Ecol. 44 (4): 336-343.

Canfield, D. E., A. N. Glazer, et al. (2010). The evolution and future of Earth's nitrogen cycle. Science. 330 (6001): 192-196.

Canfield, D. E., K. S. Habicht, et al. (2000). The Archean Sulfur Cycle and the Early History of Atmospheric Oxygen. Science. 288: 658-661.

Chan, J. M., W. Wu, et al. (2000). Construction and characterization of a heterodimeric iron protein: defining roles for adenosine triphosphate in nitrogenase catalysis. Biochemistry. 39 (24): 7221-7228.

Chan, M. K., J. Kim, et al. (1993). The nitrogenase FeMo-cofactor and P-cluster pair: 2.2 A resolution structures. Science. 260 (5109): 792-794.

Chatterjee, R., P. W. Ludden, et al. (1997). Characterization of VNFG, the delta subunit of the vnf-encoded apodinitrogenase from Azotobacter vinelandii. Implications for its role in the formation of functional dinitrogenase 2. J Biol Chem. 272 (6): 3758-3765.

Chisnell, J. R., R. Premakumar, et al. (1988). Purification of a second alternative nitrogenase from a nifHDK deletion strain of Azotobacter vinelandii. J Bacteriol. 170 (1): 27-33.

Chiu, H., J. W. Peters, et al. (2001). MgATP-Bound and nucleotide-free structures of a nitrogenase protein complex between the Leu 127 Delta-Fe-protein and the MoFe-protein. Biochemistry. 40 (3): 641-650.

Christiansen, R. L. (2000). Geology of Yellowstone Naitonal Park: The Quaternary and Pliocene Yellowstone Plateau Volcanic Field of Wyoming, Idaho, and Montana. U. S. D. o. t. Interior and U. S. G. Survey. Denver, CO, Western Region, Menlo Park, CA**:** 1-146.

Corbett, M. C., Y. Hu, et al. (2007). Conformational differences between Azotobacter vinelandii nitrogenase MoFe proteins as studied by small-angle X-ray scattering. Biochemistry. 46 (27): 8066-8074.

Curatti, L., C. S. Brown, et al. (2005). Genes required for rapid expression of nitrogenase activity in Azotobacter vinelandii. Proc Natl Acad Sci U S A. 102 (18): 6291-6296.

Curatti, L., J. A. Hernandez, et al. (2007). In vitro synthesis of the iron-molybdenum cofactor of nitrogenase from iron, sulfur, molybdenum, and homocitrate using purified proteins. Proc Natl Acad Sci U S A. 104 (45): 17626-17631.

Curatti, L., P. W. Ludden, et al. (2006). NifB-dependent in vitro synthesis of the iron-molybdenum cofactor of nitrogenase. Proc Natl Acad Sci U S A. 103 (14): 5297-5301.

Daesch, G. and L. E. Mortenson (1972). Effect of ammonia on the synthesis and function of the N 2 -fixing enzyme system in Clostridium pasteurianum. J Bacteriol. 110 (1): 103-109.

Dean, D. R., R. A. Setterquist, et al. (1990). Evidence that conserved residues Cys-62 and Cys-154 within the Azotobacter vinelandii nitrogenase MoFe protein alpha-subunit are essential for nitrogenase activity but conserved residues His-83 and Cys-88 are not. Mol Microbiol. 4 (9): 1505-1512.

Dekas, A. E., R. S. Poretsky, et al. (2009). Deep-sea archaea fix and share nitrogen in methane-consuming microbial consortia. Science. 326 (5951): 422-426.

Delwiche, C. C. and P. L. Steyn (1970). Nitrogen Isotope Fractionation in Soils and Microbial Reactions. Environ Sci Technol. 4 (11): 929-935.

DeWitt, T. J., A. Sih, et al. (1998). Costs and limits of phenotypic plasticity. Trends in Ecology & Evolution. 13: 77-81.

Dicker, H. J. and D. W. Smith (1980). Acetylene Reduction (Nitrogen Fixation) in a Delaware, USA Salt Marsh. Marine Biology. 57: 241-250.

Dilworth, M. J., D. Subramanian, et al. (1965). The adenosine triphosphate requirement for nitrogen fixation in cell-free extracts of Clostridium pasteurianum. Biochim Biophys Acta. 99 (3): 486-503.

Dingler, C., J. Kuhla, et al. (1988). Levels and activities of nitrogenase proteins in Azotobacter vinelandii grown at different dissolved oxygen concentrations. J Bacteriol. 170 (5): 2148-2152.

Dixon, R. (1998). The oxygen-responsive NIFL-NIFA complex: a novel two-component regulatory system controlling nitrogenase synthesis in gamma-proteobacteria. Arch Microbiol. 169 (5): 371-380.

Dixon, R., R. R. Eady, et al. (1980). Analysis of regulation of Klebsiella pneumoniae nitrogen fixation (nif) gene cluster with gene fusions. Nature. 286 (5769): 128-132.

Dixon, R. and D. Kahn (2004). Genetic regulation of biological nitrogen fixation. Nat Rev Microbiol. 2 (8): 621-631.

Dodsworth, J. A., N. C. Cady, et al. (2005). 2-Oxoglutarate and the PII homologues NifI1 and NifI2 regulate nitrogenase activity in cell extracts of Methanococcus maripaludis. Mol Microbiol. 56 (6): 1527-1538.

Dos Santos, P. C. and D. R. Dean (2011). Coordination and fine-tuning of nitrogen fixation in Azotobacter vinelandii. Mol Microbiol.

Dos Santos, P. C., D. R. Dean, et al. (2004). Formation and insertion of the nitrogenase iron-molybdenum cofactor. Chem Rev. 104 (2): 1159-1173.

Dos Santos, P. C., D. C. Johnson, et al. (2007). Controlled expression of nif and isc iron-sulfur protein maturation components reveals target specificity and limited functional replacement between the two systems. J Bacteriol. 189 (7): 2854-2862.

Dos Santos, P. C., A. D. Smith, et al. (2004). Iron-sulfur cluster assembly: NifU-directed activation of the nitrogenase Fe protein. J Biol Chem. 279 (19): 19705-19711.

Downie, J. A. (2005). Legume haemoglobins: symbiotic nitrogen fixation needs bloody nodules. Curr Biol. 15 (6): R196-198.

Eady, R. R. (1996). Structureminus signFunction Relationships of Alternative Nitrogenases. Chem Rev. 96 (7): 3013-3030.

Eady, R. R., R. L. Robson, et al. (1987). The vanadium nitrogenase of Azotobacter chroococcum. Purification and properties of the VFe protein. Biochem J. 244 (1): 197-207.

Eady, R. R., B. E. Smith, et al. (1972). Nitrogenase of Klebsiella pneumoniae. Purification and properties of the component proteins. Biochem J. 128 (3): 655-675.

Eigenbrode, J. L., K. H. Freeman, et al. (2008). Methylhopane biomarker hydrocarbons in Hamersley Province sediments provide evidence for Neoarchean aerobiosis. Earth and Planetary Science Letters. 273 (3-4): 323-331.

Einsle, O., F. A. Tezcan, et al. (2002). Nitrogenase MoFe-protein at 1.16 A resolution: a central ligand in the FeMo-cofactor. Science. 297 (5587): 1696-1700.

Einsle, O., F. A. Tezcan, et al. (2002). Nitrogenase MoFe-protein at 1.16 A resolution: a central ligand in the FeMo-cofactor. Science. 297 (5587): 1696-1700.

Emerich, D. W., R. V. Hageman, et al. (1981). Interactions of dinitrogenase and dinitrogenase reductase. Adv Enzymol Relat Areas Mol Biol. 52: 1-22.

Estep, M. L. F. and S. A. Macko (1984). Nitrogen isotope biogeochemistry of thermal springs. Org. Geochem. 6: 779-785.

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

Fani, R., R. Gallo, et al. (2000). Molecular evolution of nitrogen fixation: the evolutionary history of the nifD, nifK, nifE, and nifN genes. J Mol Evol. 51 (1): 1-11.

Fay, P. (1992). Oxygen relations of nitrogen fixation in cyanobacteria. Microbiol Rev. 56 (2): 340-373.

Fournier, G. P. and J. P. Gogarten (2008). Evolution of acetoclastic methanogenesis in *Methanosarcina* via horizontal gene transfer from cellulolytic *Clostridia*. J. Bacteriol. 190 (3): 1124-1127.

Fujita, Y. (1996). Protochlorophyllide reduction: a key step in the greening of plants. Plant Cell Physiol. 37 (4): 411-421.

Fujita, Y. and C. E. Bauer (2000). Reconstitution of light-independent protochlorophyllide reductase from purified bchl and BchN-BchB subunits. In vitro confirmation of nitrogenase-like features of a bacteriochlorophyll biosynthesis enzyme. J Biol Chem. 275 (31): 23583-23588.

Fujita, Y., H. Takagi, et al. (1998). Cloning of the gene encoding a protochlorophyllide reductase: the physiological significance of the co-existence of light-dependent and -independent protochlorophyllide reduction systems in the cyanobacterium Plectonema boryanum. Plant Cell Physiol. 39 (2): 177-185.

Gallon, J. R. (1981). The oxygen sensitivity of nitrogenase: a problem for biochemists and micro-organisms. Trends in Biochemical Sciences. 6: 19-23.

Garvin, J., R. Buick, et al. (2009). Isotopic evidence for an aerobic nitrogen cycle in the latest Archean. Science. 323: 1045-1048.

Gaucher, E. A., S. Govindarajan, et al. (2008). Palaeotemperature trend for Precambrian life inferred from resurrected proteins. Nature. 451 (7179): 704-707.

Gentleman, R. C., V. J. Carey, et al. (2004). Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 5 (10): R80.

Georgiadis, M. M., H. Komiya, et al. (1992). Crystallographic structure of the nitrogenase iron protein from *Azotobacter vinelandii*. Science. 257: 1653-1659.

Giulio, M. (2003). The Universal Ancestor and the Ancestor of Bacteria Were Hyperthermophiles. Journal of Molecular Evolution. 57 (6): 721 - 730.

Godfrey, L. V. and P. G. Falkowski (2009). The cycling and redox state of nitrogen in the Archaean ocean. Nature Geosci. 2: 725-729.

Goslee, S. C. and D. L. Urban (2007). The ecodist package for dissimilarity-based analysis of ecological data. J. Stat Softw. 22: 1-19.

Gribaldo, S. and C. Brochier-Armanet (2006). The origin and evolution of Archaea: a state of the art. Philos Trans R Soc Lond B Biol Sci. 361 (1470): 1007-1022.

Guindon, S. and O. Gascuel (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Systematic Biology. 52 (5): 696-704.

Hales, B. J., E. E. Case, et al. (1986). Isolation of a new vanadium-containing nitrogenase from Azotobacter vinelandii. Biochemistry. 25 (23): 7251-7255.

Hall, J. R., K. R. Mitchell, et al. (2008). Molecular characterization of the diversity and distribution of a thermal spring microbial community by using rRNA and metabolic genes. Appl Environ Microbiol. 74 (15): 4910-4922.

Hamamura, N., S. H. Olson, et al. (2005). Diversity and functional analysis of bacterial communities associated with natural hydrocarbon seeps in acidic soils at Rainbow Springs, Yellowstone National Park. Appl Environ Microbiol. 71 (10): 5943-5950.

Hamelin, J., N. Fromin, et al. (2002). nifH gene diversity in the bacterial community associated with the rhizosphere of Molinia coerulea, an oligonitrophilic perennial grass. Environ Microbiol. 4 (8): 477-481.

Hamilton, T. L., E. S. Boyd, et al. (2011). Environmental Constraints Underpin the Phylogenetic Diversity of nifH in the Yellowstone Geothermal Complex. Microb Ecol.

Hamilton, T. L., M. Ludwig, et al. (2011). Transcriptional profiling of nitrogen fixation in *Azotobacter vinelandii*. J. Bacteriol. 193: 4477-4486.

Hardy, R. W., R. D. Holsten, et al. (1968). The acetylene-ethylene assay for n(2) fixation: laboratory and field evaluation. Plant Physiol. 43 (8): 1185-1207.

Havig, J. R., J. Raymond, et al. (2010). Merging Isotopes and Community Genomics in a Siliceous Sinter-Depositing Hot Spring. Geochimica et Cosmochimica Atca.

Hearst, J. E., M. Alberti, et al. (1985). A putative nitrogenase reductase gene found in the nucleotide sequences from the photosynthetic gene cluster of *R. capsulata*. Cell. 40 (1): 219-220.

Helz, G. R., C. V. Miller, et al. (1996). Mechanism of molybdenum removal from the sea and its concentration in black shales: EXAFS evidence. Geochim. Cosmochim. Acta. 60 (19): 3631-3642.

Hernandez, J. A., L. Curatti, et al. (2008). Metal trafficking for nitrogen fixation: NifQ donates molybdenum to NifEN/NifH for the biosynthesis of the nitrogenase FeMo-cofactor. Proc Natl Acad Sci U S A. 105 (33): 11679-11684.

Hofacker, I. L. (2003). Vienna RNA secondary structure server. Nucleic Acids Res. 31 (13): 3429-3431.

Hoffman, B. M., D. R. Dean, et al. (2009). Climbing nitrogenase: toward a mechanism of enzymatic nitrogen fixation. Acc Chem Res. 42 (5): 609-619.

Holland, H. D. (2002). Volcanic gases, black smokers, and the Great Oxidation Event. . Geochim Cosmochim Acta. 66: 3811–3826.

Holland, H. D. (2006). The oxygenation of the atmosphere and oceans. Philos Trans R Soc Lond B Biol Sci. 361 (1470): 903-915.

Horner-Devine, M. C. and B. J. M. Bohannan (2006). Phylogenetic clustering and overdispersion in bacterial communities. Ecology. 87 (7): S100–S108.

Howard, J. B. and D. C. Rees (1994). Nitrogenase: a nucleotide-dependent molecular switch. Annu Rev Biochem. 63: 235-264.

Howard, J. B. and D. C. Rees (1996). *Strucutral basis of biological nitrogen fixation*. Chem Rev. 96: 2965-2982.

Howard, K. S., P. A. McLean, et al. (1986). Klebsiella pneumoniae nifM gene product is required for stabilization and activation of nitrogenase iron protein in Escherichia coli. J Biol Chem. 261 (2): 772-778.

Hu, Y., A. W. Fay, et al. (2007). P-cluster maturation on nitrogenase MoFe protein. Proc Natl Acad Sci U S A. 104 (25): 10424-10429.

Huber, H., M. J. Hohn, et al. (2002). A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. Nature. 417 (6884): 63-67.

Huelsenbeck, J. P. and F. Ronquist (2001). MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics. 17 (8): 754-755.

Huelsenbeck, J. P. and R. Ronquist (2001). MRBAYES: Bayesian inference of phylogeny. Bioinformatics. 17: 754–755.

Huelsenbeck, J. P. and R. Ronquist. (2001). MRBAYES: Bayesian inference of phylogeny. Bioinformatics. 17: 754-755.

Igarashi, R. Y. and L. C. Seefeldt (2003). Nitrogen fixation: the mechanism of the Mo-dependent nitrogenase. Crit Rev Biochem Mol Biol. 38 (4): 351-384.

Inskeep, W. P. and T. R. McDermott (2005). Geomicrobiology of acid-sulfate-chloride springs in Yellowstone National Park. Geothermal biology and geochemistry in Yellowstone National Park. W. P. Inskeep and T. R. McDermott Bozeman, MT, Montana State University. Volume 1**:** 143-162.

Jackson, C. R., H. W. Langner, et al. (2001). Molecular analysis of microbial community structure in an arsenite-oxidizing acidic thermal spring. Environ Microbiol. 3 (8): 532-542.

Jacobitz, S. and P. E. Bishop (1992). Regulation of nitrogenase-2 in Azotobacter vinelandii by ammonium, molybdenum, and vanadium. J Bacteriol. 174 (12): 3884-3888.

Jacobson, M. R., K. E. Brigle, et al. (1989). Physical and genetic map of the major nif gene cluster from Azotobacter vinelandii. J Bacteriol. 171 (2): 1017-1027.

Jacobson, M. R., V. L. Cash, et al. (1989). Biochemical and genetic analysis of the nifUSVWZM cluster from Azotobacter vinelandii. Mol Gen Genet. 219 (1-2): 49-57.

Jacobson, M. R., R. Premakumar, et al. (1986). Transcriptional regulation of nitrogen fixation by molybdenum in Azotobacter vinelandii. J Bacteriol. 167 (2): 480-486.

Javaux, E. J. (2006). Extreme life on Earth--past, present and possibly beyond. Res Microbiol. 157 (1): 37-48.

Jin, Z., M. Heinnickel, et al. (2008). Biogenesis of iron-sulfur clusters in photosystem I: holo-NfuA from the cyanobacterium Synechococcus sp. PCC 7002 rapidly and efficiently transfers [4Fe-4S] clusters to apo-PsaC in vitro. J Biol Chem. 283 (42): 28426-28435.

Joerger, R. D. and P. E. Bishop (1988). Bacterial alternative nitrogen fixation systems. Crit Rev Microbiol. 16 (1): 1-14.

Joerger, R. D. and P. E. Bishop (1988). Nucleotide sequence and genetic analysis of the nifB-nifQ region from Azotobacter vinelandii. J Bacteriol. 170 (4): 1475-1487.

Joerger, R. D., M. R. Jacobson, et al. (1989). Two nifA-like genes required for expression of alternative nitrogenases by Azotobacter vinelandii. J Bacteriol. 171 (6): 3258-3267.

Joerger, R. D., T. M. Loveless, et al. (1990). Nucleotide sequences and mutational analysis of the structural genes for nitrogenase 2 of Azotobacter vinelandii. J Bacteriol. 172 (6): 3400-3408.

Joerger, R. D., R. Premakumar, et al. (1986). Tn5-induced mutants of Azotobacter vinelandii affected in nitrogen fixation under Mo-deficient and Mo-sufficient conditions. J Bacteriol. 168 (2): 673-682.

Joerger, R. D., E. D. Wolfinger, et al. (1991). The gene encoding dinitrogenase reductase 2 is required for expression of the second alternative nitrogenase from Azotobacter vinelandii. J Bacteriol. 173 (14): 4440-4446.

Johnson, D. C., D. R. Dean, et al. (2005). Structure, function, and formation of biological iron-sulfur clusters. Annu Rev Biochem. 74: 247-281.

Johnson, D. C., P. C. Dos Santos, et al. (2005). NifU and NifS are required for the maturation of nitrogenase and cannot replace the function of isc-gene products in Azotobacter vinelandii. Biochem Soc Trans. 33 (Pt 1): 90-93.

Johnson, J. B. and K. S. Omland (2004). Model selection in ecology and evolution. Trends Ecol Evol. 19 (2): 101-108.

Kaiser, J. T., Y. Hu, et al. (2011). Structure of precursor-bound NifEN: a nitrogenase FeMo cofactor maturase/insertase. Science. 331 (6013): 91-94.

Kaminski, P. A., F. Norel, et al. (1988). Characterization of the fixABC region of Azorhizobium caulinodans ORS571 and identification of a new nitrogen fixation gene. Mol Gen Genet. 214 (3): 496-502.

Kassen, R. and P. B. Rainey (2004). The ecology and genetics of microbial diversity. Annu Rev Microbiol. 58: 207-231.

Kasting, J. F. (1981). Limits on Oxygen Concentration in the Prebiological Atmosphere and the Rate of Abiotic Fixation of Nitrogen. Journal of Geophysical Research. 86: 1147 - 1158.

Kasting, J. F. (2001). Earth history. The rise of atmospheric oxygen. Science. 293 (5531): 819-820.

Kasting, J. F. and J. L. Siefert (2001). Biogeochemistry: The nitrogen fix. Nature. 412: 26 - 27.

Kaufman, A. J., D. T. Johnston, et al. (2007). Late Archean biospheric oxygenation and atmospheric evolution. Science. 317 (5846): 1900-1903.

Kendall, B., C. T. Reinhard, et al. (2010). Pervasive oxygenation along late Archaean ocean margins. Nature Geosci. 3 (9): 647-652.

Kennedy, C., and D. Dean. 1992. The nifU, nifS and nifV gene products are required for activity of all three nitrogenases of Azotobacter vinelandii. Mol Gen Genet 231:494-8. and D. Dean (1992). The *nifU*, *nifS* and *nifV* gene products are required for activity of all three nitrogenases of *Azotobacter vinelandii*. Mol Gen Genet. 231: 494-498.

Kennedy, C., P. Rudnick, et al. (2005). Genus Azotobacter. Bergey's manual of sytematic bacteriology. G. M. Garrity. New York, NY, Spring-Verlag. 2, part B**:** 384-401.

Kennedy, C. and A. Toukdarian (1987). Genetics of azotobacters: applications to nitrogen fixation and related aspects of metabolism. Annu Rev Microbiol. 41: 227-258.

Kent, H. M., M. Baines, et al. (1990). Analysis of site-directed mutations in the alpha- and beta-subunits of Klebsiella pneumoniae nitrogenase. Mol Microbiol. 4 (9): 1497-1504.

Kent, H. M., I. Ioannidis, et al. (1989). Site-directed mutagenesis of the Klebsiella pneumoniae nitrogenase. Effects of modifying conserved cysteine residues in the alpha- and beta-subunits. Biochem J. 264 (1): 257-264.

Kertesz, M., Y. Wan, et al. (2003). Genome-wide measurement of RNA secondary structure in yeast. Nature. 467 (7311): 103-107.

Kessler, P. S., C. Blank, et al. (1998). The nif gene operon of the methanogenic archaeon Methanococcus maripaludis. J Bacteriol. 180 (6): 1504-1511.

Kessler, P. S., J. McLarnan, et al. (1997). Nitrogenase phylogeny and the molybdenum dependence of nitrogen fixation in Methanococcus maripaludis. J Bacteriol. 179 (2): 541-543.

Kim, J. and D. C. Rees (1992). Crystallographic structure and functional implications of the nitrogenase molybdenum-iron protein from *Azotobacter vinelandii*. Nature. 360: 553-560.

Kim, J. and D. C. Rees (1992). Structural models for the metal centers in the nitrogenase molybdenum-iron protein. Science. 257 (5077): 1677-1682.

Kim, J., D. Woo, et al. (1993). X-ray crystal structure of the nitrogenase molybdenum-iron protein from Clostridium pasteurianum at 3.0-A resolution. Biochemistry. 32 (28): 7104-7115.

Knowles, R. (1979). Denitrification, acetylene reduction, and methane metabolism in lake sediment exposed to acetylene. Appl Environ Microbiol. 38 (3): 486-493.

Koonin, E. V. (1993). A superfamily of ATPases with diverse functions containing either classical or deviant ATP-binding motif. J Mol Biol. 229 (4): 1165-1174.

Krol, A. J., J. G. Hontelez, et al. (1982). On the operon structure of the nitrogenase genes of Rhizobium leguminosarum and Azotobacter vinelandii. Nucleic Acids Res. 10 (14): 4147-4157.

Laguerre, G., P. van Berkum, et al. (1997). Genetic diversity of rhizobial symbionts isolated from legume species within the genera Astragalus, Oxytropis, and Onobrychis. Appl Environ Microbiol. 63 (12): 4748-4758.

Langlois, R. J., J. LaRoche, et al. (2005). Diazotrophic diversity and distribution in the tropical and subtropical Atlantic Ocean. Appl Environ Microbiol. 71 (12): 7910-7919.

Langner, H. W., C. R. Jackson, et al. (2001). Rapid oxidation of arsenite in a hot spring ecosystem, Yellowstone National Park. Environ Sci Technol. 35 (16): 3302-3309.

Lanzilotta, W. N., M. J. Ryle, et al. (1995). Nucleotide hydrolysis and protein conformational changes in Azotobacter vinelandii nitrogenase iron protein: defining the function of aspartate 129. Biochemistry. 34 (34): 10713-10723.

Lanzilotta, W. N. and L. C. Seefeldt (1997). Changes in the midpoint potentials of the nitrogenase metal centers as a result of iron protein-molybdenum-iron protein complex formation. Biochemistry. 36 (42): 12976-12983.

Larkin, M. A., G. Blackshields, et al. (2007). Clustal W and Clustal X version 2.0. Bioinformatics. 23: 2947-2948.

Lee, S. C. and R. H. Holm (2004). The clusters of nitrogenase: synthetic methodology in the construction of weak-field clusters. Chem Rev. 104 (2): 1135-1158.

Leigh, J. A. (2000). Nitrogen fixation in methanogens: the archaeal perspective. Curr Issues Mol Biol. 2 (4): 125-131.

Lindahl, P. A., E. P. Day, et al. (1985). Mossbauer, EPR, and magnetization studies of the Azotobacter vinelandii Fe protein. Evidence for a [4Fe-4S]1+ cluster with spin S = 3/2. J Biol Chem. 260 (20): 11160-11173.

Lindahl, P. A., N. J. Gorelick, et al. (1987). EPR and Mossbauer studies of nucleotide-bound nitrogenase iron protein from Azotobacter vinelandii. J Biol Chem. 262 (31): 14945-14953.

Locher, K. P., M. Hans, et al. (2001). Crystal structure of the Acidaminococcus fermentans 2-hydroxyglutaryl-CoA dehydratase component A. J Mol Biol. 307 (1): 297-308.

Lovley, D. R. and M. J. Klug (1983). Sulfate reducers can outcompete methanogens at freshwater sulfate concentrations. Appl Environ Microbiol. 45: 187-192.

Lowe, D. J., K. Fisher, et al. (1993). Klebsiella pneumoniae nitrogenase: pre-steady-state absorbance changes show that redox changes occur in the MoFe protein that depend on substrate and component protein ratio; a role for P-centres in reducing dinitrogen? Biochem J. 292 (Pt 1): 93-98.

Lowenstern, J. B., R. B. Smith, et al. (2006). Monitoring super-volcanoes: geophysical and geochemical signals at Yellowstone and other large caldera systems. Philos Transact A Math Phys Eng Sci. 364 (1845): 2055-2072.

Lowery, R. G., L. L. Saari, et al. (1986). Reversible regulation of the nitrogenase iron protein from Rhodospirillum rubrum by ADP-ribosylation in vitro. J Bacteriol. 166 (2): 513-518.

Lozupone, C. A. and R. Rob Knight (2008). Species divergence and the measurement of microbial diversity. FEMS Microbiol Rev. 32: 557–578.

Ludwig, M. and D. A. Bryant (2011). Transcription Profiling of the Model Cyanobacterium Synechococcus sp. Strain PCC 7002 by Next-Gen (SOLiD) Sequencing of cDNA. Front Microbiol. 2: 41.

Macur, R. E., H. W. Langner, et al. (2004). Linking geochemical processes with microbial community analysis: successional dynamics in an arsenic-rich, acid-sulphate-chloride geothermal spring. Geobiology. 2: 163–177.

Magnuson, J. K., T. D. Paustian, et al. (1997). Nitrogenase iron-molybdenum cofactor binding site: Protein conformational changes associated with cofactor binding. Tetrahedron. 53: 11971-11984.

Maier, R. J. and F. Moshiri (2000). Role of the Azotobacter vinelandii nitrogenase-protective shethna protein in preventing oxygen-mediated cell death. J Bacteriol. 182 (13): 3854-3857.

Martinez-Argudo, I., R. Little, et al. (2005). Nitrogen fixation: key genetic regulatory mechanisms. Biochem Soc Trans. 33 (Pt 1): 152-156.

Mayer, S. M., C. A. Gormal, et al. (2002). Crystallographic analysis of the MoFe protein of nitrogenase from a nifV mutant of Klebsiella pneumoniae identifies citrate as a ligand to the molybdenum of iron molybdenum cofactor (FeMoco). J. Biol. Chem. 277: 35263-35266.

Mayer, S. M., D. M. Lawson, et al. (1999). New insights into structure-function relationships in nitrogenase: A 1.6 A resolution X-ray crystallographic study of Klebsiella pneumoniae MoFe-protein. J Mol Biol. 292 (4): 871-891.

McCann, K. S. (2000). The diversity-stability debate. Nature. 405 (6783): 228-233.

McMaster, G. K. and G. G. Carmichael (1977). Analysis of single- and double-stranded nucleic acids on polyacrylamide and agarose gels by using glyoxal and acridine orange. Proc Natl Acad Sci U S A. 74 (11): 4835-4838.

Mehta, M. P. and J. A. Baross (2006). Nitrogen fixation at 92 degrees C by a hydrothermal vent archaeon. Science. 314 (5806): 1783-1786.

Mehta, M. P., D. A. Butterfield, et al. (2003). Phylogenetic diversity of nitrogenase (nifH) genes in deep-sea and hydrothermal vent environments of the Juan de Fuca Ridge. Appl Environ Microbiol. 69 (2): 960-970.

Meyer, J. (1981). Comparison of carbon monoxide, nitric oxide, and nitrite as inhibitors of the nitrogenase from Clostridium pasteurianum. Arch Biochem Biophys. 210 (1): 246-256.

Michaelis, W., R. Seifert, et al. (2002). Microbial reefs in the Black Sea fueled by anaerobic oxidation of methane. Science. 297 (5583): 1013-1015.

Moisander, P. H., A. E. Morrison, et al. (2007). Spatial-temporal variability in diazotroph assemblages in Chesapeake Bay using an oligonucleotide *nifH* microarray. Environ. Microbiol. 9 (7): 1823–1835.

Mojzsis, S. J., G. Arrhenius, et al. (1996). Evidence for life on Earth before 3,800 million years ago. Nature. 384 (6604): 55-59.

Morgan, T. V., R. C. Prince, et al. (1986). Electrochemical titration of the S=3/2 and S=1/2 states of the iron of nitrogenase. FEBS LETTERS. 206: 4-8.

Mouncey, N. J., L. A. Mitchenall, et al. (1995). Mutational analysis of genes of the mod locus involved in molybdenum transport, homeostasis, and processing in Azotobacter vinelandii. J Bacteriol. 177 (18): 5294-5302.

Muraki, N., J. Nomata, et al. (2010). X-ray crystal structure of the light-independent protochlorophyllide reductase. Nature. 465 (7294): 110-114.

Navarro-González, R., C. P. McKay, et al. (2001). A possible nitrogen crisis for Archaean life due to reduced nitrogen fixation by lightning. Nature. 412: 61-64.

Nielsen, M., C. Lundegaard, et al. (2010). CPHmodels-3.0-remote homology modeling using structure-guided sequence profiles. Nucleic Acids Res. 38 (suppl 2): W576-W581.

Nordstrom, D. K., J. W. Ball, et al. (2005). Ground water to surface water: chemistry of thermal outflows in Yellowstone National Park. Geothermal biology and geochemistry in Yellowstone National Park. W. P. Inskeep and T. R. McDermott. Bozeman, Montana State University**:** 143-162.

Norris, P. R., J. C. Murrell, et al. (1995). The potential for diazotrophy in iron-and sulfur-oxidizing acidophilic bacteria. Arch. Microbiol. 164: 294-300.

Oda, Y., S. K. Samanta, et al. (2005). Functional genomic analysis of three nitrogenase isozymes in the photosynthetic bacterium Rhodopseudomonas palustris. J Bacteriol. 187 (22): 7784-7794.

Orme-Johnson, W. H., W. D. Hamilton, et al. (1972). Electron paramagnetic resonance of nitrogenase and nitrogenase components from Clostridium pasteurianum W5 and Azotobacter vinelandii OP. Proc Natl Acad Sci U S A. 69 (11): 3142-3145.

Page, R. D. M. (1996). TREEVIEW: An application to display phylogenetic trees on personal computers. Comp. Appl. Biosci. 12: 357-358.

Pagel, M. (1999). Inferring the historical patterns of biological evolution. Nature. 401: 877-884.

Parsons, I., M. R. Lee, et al. (1998). Biochemical evolution II: Origin of life in tubular microstructures on weathered feldspar surfaces. Proc. Natl. Acad. Sci. USA. 95: 15173–15176.

Penny, D. and A. Poole (1999). The nature of the last universal common ancestor. Curr. Opin. Gen. Dev. 9: 672–677.

Peters, J. W., K. Fisher, et al. (1995). Nitrogenase structure and function: a biochemical-genetic perspective. Annu. Rev. Microbiol. 49: 335-366.

Peters, J. W., M. H. Stowell, et al. (1997). Redox-dependent structural changes in the nitrogenase P-cluster. Biochemistry. 36 (6): 1181-1187.

Peters, J. W. and R. K. Szilagyi (2006). Exploring new frontiers of nitrogenase structure and mechanism. Curr Opin Chem Biol. 10 (2): 101-108.

Pilcher, C. B. (2003). Biosignatures of early earths. Astrobiology. 3 (3): 471-486.

Poly, F., L. Ranjard, et al. (2001). Comparison of nifH gene pools in soils and soil microenvironments with contrasting properties. Appl Environ Microbiol. 67 (5): 2255-2262.

Poole, R. K. and S. Hill (1997). Respiratory protection of nitrogenase activity in Azotobacter vinelandii--roles of the terminal oxidases. Biosci Rep. 17 (3): 303-317.

Posada, D. (2006). ModelTest Server: a web-based tool for the statistical selection of models of nucleotide substitution online. Nucleic Acids Res. 34 (Web Server issue): W700-703.

Poulton, S. W., P. Fralick, et al. (2010). Spatial variability in oceanic redox structure 1.8 billion years ago. Nat. Geosci. 3: 486-490.

Premakumar, R., M. R. Jacobson, et al. (1992). Characterization of transcripts expressed from nitrogenase-3 structural genes of Azotobacter vinelandii. Can J Microbiol. 38 (9): 929-936.

Premakumar, R., R. N. Pau, et al. (1998). Regulation of the transcriptional activators AnfA and VnfA by metals and ammonium in Azotobacter vinelandii. FEMS Microbiol Lett. 164 (1): 63-68.

Purcell, D., U. Sompong, et al. (2007). The effects of temperature, pH and sulphide on the community structure of hyperthermophilic streamers in hot springs of northern Thailand. FEMS Microbiol. Ecol. 60 (3): 456 - 466.

Rangaraj, P. and P. W. Ludden (2002). Accumulation of 99Mo-containing iron-molybdenum cofactor precursors of nitrogenase on NifNE, NifH, and NifX of Azotobacter vinelandii. J Biol Chem. 277 (42): 40106-40111.

Raymond, J., J. L. Siefert, et al. (2004). The natural history of nitrogen fixation. Mol Biol Evol. 21 (3): 541-554.

Rees, D. C. and J. B. Howard (2000). Nitrogenase: standing at the crossroads. Curr Opin Chem Biol. 4 (5): 559-566.

Reinhard, C. T., R. Raiswell, et al. (2009). A late Archean sulfidic sea stimulated by early oxidative weathering of the continents. Science. 326 (5953): 713-716.

Reiter, B., H. Burgmann, et al. (2003). Endophytic nifH gene diversity in African sweet potato. Can J Microbiol. 49 (9): 549-555.

Reysenbach, A. L. and S. L. Cady (2001). Microbiology of ancient and modern hydrothermal systems. Trends Microbiol. 9 (2): 79-86.

Reysenbach, A. L. and E. Shock (2002). Merging genomes with geochemistry in hydrothermal ecosystems. Science. 296 (5570): 1077-1082.

Robinson, A. C., B. K. Burgess, et al. (1986). Activity, reconstitution, and accumulation of nitrogenase components in Azotobacter vinelandii mutant strains containing defined deletions within the nitrogenase structural gene cluster. J Bacteriol. 166 (1): 180-186.

Robinson, A. C., D. R. Dean, et al. (1987). Iron-molybdenum cofactor biosynthesis in Azotobacter vinelandii requires the iron protein of nitrogenase. J Biol Chem. 262 (29): 14327-14332.

Robson, R. L., R. R. Eady, et al. (1986). The alternative nitrogenase of *Azotobacter chroococcum* is a vanadium enzyme. Nature. 322: 388-390.

Rodriguez-Trelles, F., R. Tarrio, et al. (2002). A methodological bias toward overestimation of molecular evolutionary time scales. Proc. Natl. Acad. Sci. USA. 12: 8112-8115.

Ronquist, F. and J. P. Huelsenbeck (2003). MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 19: 1572-1574.

Rubio, L. M. and P. W. Ludden (2008). Biosynthesis of the iron-molybdenum cofactor of nitrogenase. Annu Rev Microbiol. 62: 93-111.

Rubio, L. M., P. Rangaraj, et al. (2002). Cloning and mutational analysis of the gamma gene from Azotobacter vinelandii defines a new family of proteins capable of metallocluster binding and protein stabilization. J Biol Chem. 277 (16): 14299-14305.

Ruttimann-Johnson, C., L. M. Rubio, et al. (2003). VnfY is required for full activity of the vanadium-containing dinitrogenase in Azotobacter vinelandii. J Bacteriol. 185 (7): 2383-2386.

Ryle, M. J., W. N. Lanzilotta, et al. (1995). Evidence for a central role of lysine 15 of Azotobacter vinelandii nitrogenase iron protein in nucleotide binding and protein conformational changes. J Biol Chem. 270 (22): 13112-13117.

Ryle, M. J., W. N. Lanzilotta, et al. (1996). Elucidating the mechanism of nucleotide-dependent changes in the redox potential of the [4Fe-4S] cluster in nitrogenase iron protein: the role of phenylalanine 135. Biochemistry. 35 (29): 9424-9434.

Ryle, M. J., W. N. Lanzilotta, et al. (1996). Circular dichroism and x-ray spectroscopies of Azotobacter vinelandii nitrogenase iron protein. MgATP and MgADP induced protein conformational changes affecting the [4Fe-4S] cluster and characterization of a [2Fe-2S] form. J Biol Chem. 271 (3): 1551-1557.

Sanderson, M. J. (1997). A nonparametric approach to estimating divergence times in the absence of rate constancy. . Mol. Biol. Evol. 14: 1218–1231.

Sanderson, M. J. (2002). Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. . Mol. Biol. Evol. 19: 101-109.

Sanderson, M. J. (2004). r8s, version 1.70. User's manual.

Sarma, R., B. M. Barney, et al. (2008). Crystal structure of the L protein of Rhodobacter sphaeroides light-independent protochlorophyllide reductase with MgADP bound: a homologue of the nitrogenase Fe protein. Biochemistry. 47 (49): 13004-13015.

Schindelin, H., C. Kisker, et al. (1997). Structure of ADP x AIF4(-)-stabilized nitrogenase complex and its implications for signal transduction. Nature. 387 (6631): 370-376.

Schlessman, J. L., D. Woo, et al. (1998). Conformational variability in structures of the nitrogenase iron proteins from Azotobacter vinelandii and Clostridium pasteurianum. J Mol Biol. 280 (4): 669-685.

Schloss, P. D. and J. Handelsman (2005). Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. Appl Environ Microbiol. 71 (3): 1501-1506.

Schmehl, M., A. Jahn, et al. (1993). Identification of a new class of nitrogen fixation genes in Rhodobacter capsulatus: a putative membrane complex involved in electron transport to nitrogenase. Mol Gen Genet. 241 (5-6): 602-615.

Schmid, B., O. Einsle, et al. (2002). Biochemical and structural characterization of the cross-linked complex of nitrogenase: comparison to the ADP-AlF4(-)-stabilized structure. Biochemistry. 41 (52): 15557-15565.

Schneider, K., U. Gollan, et al. (1997). Comparative biochemical characterization of the iron-only nitrogenase and the molybdenum nitrogenase from Rhodobacter capsulatus. Eur J Biochem. 244 (3): 789-800.

Schneider, K., U. Gollan, et al. (1994). Rapid purification of the protein components of a highly active "iron only" nitrogenase. Naturwissenschaften. 81 (9): 405-408.

Schneider, K., A. Muller, et al. (1991). Demonstration of a molybdenum- and vanadium-independent nitrogenase in a nifHDK-deletion mutant of Rhodobacter capsulatus. Eur J Biochem. 195 (3): 653-661.

Schoonen, M. A. and Y. Xu (2001). Nitrogen reduction under hydrothermal vent conditions: implications for the prebiotic synthesis of C-H-O-N compounds. Astrobiology. 1 (2): 133-142.

Schopf, J. W. (1992). Microfossils of the early Archean Apex Chert: New evidence of the antiquity of life. Science. 260: 640-646.

Schulze, G. E. (1992). Binding of nucleotides by proteins: Current opinion in structural biology Current Opinion in Structural Biology. 2: 61 - 67.

Scott, C., T. W. Lyons, et al. (2008). Tracing stepwise oxygenation of the Proterozoic ocean. Nature. 452: 456-459.

Scott, D. J., D. R. Dean, et al. (1992). Nitrogenase-catalyzed ethane production and CO-sensitive hydrogen evolution from MoFe proteins having amino acid substitutions in an alpha-subunit FeMo cofactor-binding domain. J Biol Chem. 267 (28): 20002-20010.

Scott, D. J., H. D. May, et al. (1990). Role for the nitrogenase MoFe protein alpha-subunit in FeMo-cofactor binding and catalysis. Nature. 343 (6254): 188-190.

Seefeldt, L. C. and D. J. Arp (1986). Purification to homogeneity of Azotobacter vinelandii hydrogenase: a nickel and iron containing alpha beta dimer. Biochimie. 68 (1): 25-34.