South Dakota State University
Open PRAIRIE: Open Public Research Access Institutional Repository and Information Exchange

Electronic Theses and Dissertations

2021

Free-living Diazotrophs and the Nitrogen Cycle in Natural Grassland Revealed by Culture Dependent and Independent Approaches

Amrit Koirala

Follow this and additional works at: https://openprairie.sdstate.edu/etd2

Part of the Microbiology Commons

FREE-LIVING DIAZOTROPHS AND THE NITROGEN CYCLE IN NATURAL GRASSLAND REVEALED BY CULTURE DEPENDENT AND INDEPENDENT APPROACHES

BY

AMRIT KOIRALA

A dissertation submitted in partial fulfillment of the requirements for the

Doctor of Philosophy

Major in Biological Sciences

Specialization in Microbiology

South Dakota State University

2021

FREE-LIVING DIAZOTROPHS AND THE NITROGEN CYCLE IN NATURAL GRASSLAND REVEALED BY CULTURE DEPENDENT AND INDEPENDENT APPROACHES

AMRIT KOIRALA

This dissertation is approved as a creditable and independent investigation by a candidate for the Doctor of Philosophy in Biological Sciences and is acceptable for meeting the dissertation requirements for this degree. Acceptance of this does not imply that the conclusions reached by the candidate are necessarily the conclusions of the major department.

Volker S. Brözel, PhD.
Dissertation Advisor — Date

Radhey S. Kaushik, PhD.
Head, Department of Biology and Microbiology — Date

Director, Graduate School — Date

ACKNOWLEDGEMENTS

I would like to express my immense gratitude to all the individuals who directly or indirectly contributed to forge this work. I am deeply indebted to my academic advisor Dr. Volker S. Brözel for providing me with this gratifying opportunity to pursue my PhD and this work would not have been completed without his acuity, guidance, and inspiration throughout my research.

I would also like to thank all my committee members, Dr. Senthil Subramanian, Dr. Madhav Nepal, Dr. Joy Scaria, Dr. David Davis and Dr. Shin-Yi Marzano for their valuable insights, suggestions, and time to discuss my questions whenever I needed. In addition, I am thankful to my lab mate Bibha Dahal, Gitanjali Nanda Kafle, Nabilah Ali Alshibli, and Bikarm Das for helping me to learn basics in the lab and collaboration in some of my projects. I cannot forget to thank all the staff at Department of Biology and Microbiology.

Last but not the least, my special gratitude goes to my parents for their perpetual moral support and my wife for taking care of me and beautiful daughters when I was working on my project.

CONTENTS

ABBREVIATIONS ix

LIST OF FIGURES xi

LIST OF TABLES xx

ABSTRACT xxi

Chapter 1: Literature Review 1

1.1 Introduction 1

1.1.1 Global nitrogen cycle 3

1.1.2 Biological Nitrogen Fixation (BNF) 10

1.1.3 Nitrogenase complex 11

1.1.4 Diversity of diazotrophs 27

1.1.5 Approaches to study diversity of diazotrophs 29

1.2 Prairie field and nitrogen fixation in grassland ecosystem 31

1.3 Research questions and specific objectives 33

Chapter 2: Phylogeny of Nitrogenase Structural and Assembly Components Reveals New Insights into the Origin and Distribution of Nitrogen Fixation across Bacteria and Archaea 35

2.1 Introduction 35

2.2 Materials and Methods 39

2.3 Results 41

2.3.1. Biochemical Evidence of Diazotrophy 44

2.3.2. Phylogeny of Nitrogenase 45

2.3.3. Distribution and Phylogeny of Nitrogenase in Various Phyla 46

2.3.4. Alternative Nitrogenases 58

2.3.5. Horizontal Gene Transfer (HGT) 59

2.4 Discussion... 61

2.5 Conclusion 68

Chapter 3: Study of Diversity of Free-living Nitrogen Fixing Bacteria in Sioux Prairie by Culture-based Approach. 70

3.1 Introduction.. 70

3.2 Materials and methods 74

3.2.1 Sampling site 74

3.2.2 Isolation of potential diazotrophic bacteria 75

3.2.3 Purification of isolates to obtain pure cultures of potential diazotrophs 76

3.2.4 Extraction of genomic DNA 78

3.2.5 Amplification of 16S rRNA sequence and classification of isolates............. 79

3.2.6 Amplification and classification of *nifH* sequences. 79

3.2.7 Quantification of nitrogen fixation by selected isolates 81

3.3 Results 81

3.3.1 Chemical analysis of NFM.. 81

3.3.2 Enumeration and colony characteristics of putative diazotrophs 83

3.3.3 Diversity of potential diazotrophs 83

3.3.4 Survey of *nifH* gene among the isolates 89

3.3.5 Quantification of nitrogen fixation by $^{15}N_2$ assimilation 92

3.3.6 Quantification of nitrogen fixation by ARA 93

3.4 Discussion 93

3.5 Conclusion 98

Chapter 4: Characterization of *Streptomyces* Isolate Growing in Nitrogen Free Medium and Study of Differentially Expressed Genes in Nitrogen Deplete and Replete Condition 99

4.1 Introduction 99

4.2 Materials and methods 101

4.2.1 Bacterial specimen 101

4.2.2 Growth in nitrogen free medium 102

4.2.3 Microscopy 102

4.2.4 Growth rate measurement 103

4.2.5 Extraction and sequencing of complete genome 103

4.2.6 Extraction and sequencing of total RNA 103

4.2.7 Processing and analysis of sequencing data 104

4.3 Results 105

4.3.1 Culture characteristics 105

4.3.2 Microscopy 106

4.3.3 Growth characteristics 107

4.3.4 Measurement of nitrogen fixation 109

4.3.5 Analysis of *Streptomyces* genome 109

4.3.6 Differential gene expression 112

4.4 Discussion 119

4.5 Conclusions 124

Chapter 5: Diversity and Expression of the Nitrogen Cycling Genes in Sioux Prairie Soil Assessed by Meta-transcriptomic Approach. 126

5.1 Introduction 126

5.2 Materials and methods 130

5.2.1 Sampling site 130

5.2.2 Extraction of total RNA 130

5.2.3 RNA Quality check and HTS 131

5.2.4 Quality check and processing of sequence data 131

5.3 Results 132

5.3.1 Quantification and quality of RNA extract 132

5.3.2 Sequencing quality 133

5.3.3 De novo assembly of transcriptome 134

5.3.4 Diversity of the prokaryotes 134

5.3.5 N_2 cycle 135

5.4 Discussion 146

5.5 Conclusion 150

References 152

Appendix 180

ABBREVIATIONS

ADP: Adenosine Di-Phosphate

AEC: Adenylate Energy Change

ARA: Acetylene Reduction Assay

ATP: Adenosine Tri- Phosphate

BSC: Biological Soil Crust

BLAST: Basic Local Alignment Search Tool

BNF: Biological Nitrogen Fixation

bp: base pair

BSC: Biological Soil Crust

CDS: Coding Sequence

CFU: Colony Forming Unit

CPR: Candidate Phylum Radiation

Da: Dalton

DNA: Deoxyribonucleic acid

dNTP: Deoxyribonucleic triphosphate

HGT: Horizontal Gene Transfer

HPC: High Performance Computing

HTS: High Throughput Sequencing

ITS: Internal Transcribed Spacer

NCBI: National Center for Biotechnology Information

NFA: Nitrogen Free Agar

NFM: Nitrogen Free Medium

NGS: Next Generation Sequencing

NR: Nitrate Reductase

OTU: Operational Taxonomical Unit

PVC: Planctomyces, Verrucomicrobia, and Chlamydia

RNA: Ribonucleic Acid

TOL: Tree of Life

UCYN: Uncultured Unicellular Cyanobacterium

LIST OF FIGURES

Figure 1. 1 Oxidation states of nitrogen in various stages in the global nitrogen cycle. Dinitrogen and Nitrous oxide with oxidation state zero and 1 respectively are non-fixed forms and not accessible to the majority of living organisms. The most reduced ammonium (-3) and the most oxidized nitrate (+5) are most easily assimilated by most of the organims and other radicals are formed in various stages in nitrogen cycle (Zhang et al., 2020). 3

Figure 1. 2 Microbial nitrogen cycle highlighting the input process, internal cycling and output process and the microbial enzymes involved in each step. Adapted from Zhang et al. 2020 and the KEGG nitrogen metabolism map (Kanehisa & Goto, 2000; Zhang et al., 2020). 7

Figure 1. 3 Structure of one catalytic half of Molybdenum nitrogenase. (A) Component II (Fe protein homodimer) shown in tan, and Component I (dinitrogenase) shown in green (α subunit) and cyan (β subunit). (B) Fe protein has one Fe-S cluster and MoFe protein has P-cluster and a MoFe co (M-cluster). The M cluster has an inorganic [7Fe-9S-C-Mo] cluster and a molecule of R-homocitrate (Hoffman et al., 2014). 13

Figure 1. 4 Fe protein cycle showing the multistep reduction of MoFe protein by Fe protein. The reduced Fe protein $(Fe(red)(ATP)_2$ forms the complex with MoFe protein (MoFe(n): where n= number of electrons in the active site MoFe protein). During the dissociation of the complex Fe protein is oxidized $(Fe(ox)ADP_2)$ and MoFe protein gains one more electron (MoFe(n+1)). The oxidized Fe protein is cycled again by reduction by nucleotide exchange and the reduced MoFe(n+1) can either bind to substrate or keep on gaining electrons by forming the complex with recycled Fe protein. From (Newton, 2015). 17

Figure 1. 5 MoFe protein cycle as depicited by the Lowe and Thorneley model. MoFe protein is cycled between eight oxidation states (E_0 to E_7). Different substrates of nitrogenase bind reversibly to differnet MoFe protein redox states. For example N_2 binds to the E_3 or E_4 state and C_2H_2 binds to either the E_1 or E_2 state. From (Hoffman et al., 2014) 18

Figure 1. 6 Genomic arrangement of nitrogen fixing genes in A. vinelandii. The nif genes occur in two clusters and A. vinelandii has separate anf and vnf cluster as well. From (Ribbe, 2011) 19

Figure 1. 7 The NifL and NifA system acts as regulator of nif gene expression. NifA is the activator of RNA polymerase σ^{54} factor and NifL is the anti-activator of NifA. Both of these proteins have multiple domains to receive the signal about redox state, C:N ratio and ATP level. From (Poza-Carrión et al., 2015) 21

Figure 1. 8 Phylogeny of nitrogenase and homologous using NifHD protein. It shows five groups of which functional nitrogenase were found only in first three groups. Group IV contains uncharacterized Nif proteins and Group V contains protochlorophyllide reductase proteins.From (Raymond et al., 2004). 23

Figure 1. 9 NifH phylogeny given by Zehr et al. It has four clusters which show loose association with 16S rRNA phylogeny. From (Zehr et al., 2003). 26

Figure 2. 1 Schematic representation of FeMo-co biosynthesis. NifB (**a**) is a homodimer and belongs to the SAM radical protein family. NifB supplies homocitrate and Mo free, Fe-S cluster (precursor of Fe-Mo cofactor) to the Apo-NifEN heterotetramer (**b**). NifEN is homologous to NifDK and acts as molecular scaffold for assembly of Fe-Mo cofactor along with NifH (**c**,**d**). The assembled Fe-Mo cofactor is transferred to the Nitrogenase apoenzyme (**e**), which forms the nitrogenase holoenzyme (**f**) along with NifH. NifH has a

role in both assembly of the Fe-Mo cofactor as well as reduction of dinitrogen (Burén et al., 2020). 39

Figure 2. 2 Distribution, habitat, and lifestyle of diazotrophs in various bacterial and Archaeal phyla. (**a**) The black bar represents the total number of genomes sequenced, and the gray bar the total number of diazotrophs in that phylum. Labels on top of black bar indicate the total number of sequenced genomes. Labels in the gray bar represent the number of genomes containing nifHDKENB in that phylum (numbers in parentheses are the percentage of sequenced genomes that have nifHDKENB). (**b**) Distribution of organisms having all six nif genes by habitat and (**c**) pie diagram showing the lifestyle adopted by various diazotrophs. 42

Figure 2. 3 Distribution of diazotrophs by genus. (**a**) Distribution of diazotrophs in the tree of life proposed by Zhu et al. [23]. Nexus tree file for all prokaryotes was visualized in iTOL and leaves were annotated by genus according to the evidence of diazotrophy observed in this study. Solid and open symbols represent genera with and without biochemical evidence respectively. Genera without a star had neither genomic nor biochemical evidence of diazotrophy. (**b**) Distribution of the diazotrophic genera by the number of species observed in each genus. Bubble size corresponds to the number of assemblies of that genus and phylum....................43

Figure 2. 4 Molecular phylogenetic analysis of concatenated NifHDKENB proteins by FastTree using the JTT+CAT evolution model. Each clade is highlighted by the bacterial or archaeal phylum and Proteobacteria are further divided into classes. Inset: Phylogenetic analysis of the NifHDK protein sequences with concatenated bclLNB as outgroup. All the clusters are labeled as suggested by Raymond (Raymond et al., 2004). 47

Figure 2. 5 NifHDKENB Clusters IA and IB showing the phylogenetic relationship among Proteobacterias obtained by FastTree. 48

Figure 2. 6 NifHDKENB phylogeny highlighting nitrogenase distribution in Archaea (**a**), Firmicutes (**b**), and overall distribution of Cluster IIC (**c**), obtained by FastTree, and manually rooted based on the position of the outgroup (Cluster V) in the NifHDK tree. .52

Figure 2. 7 Phylogenetic relationship among Cyanobacteria in Cluster IC of the Concatenated NifHDKENB tree obtained by FastTree. 56

Figure 2. 8 Cluster III of the concatenated AnfHDK/VnfHDK tree obtained by FASTTREE, showing the phylogenetic relationship between isolates carrying alternative forms of nitrogenase. 59

Figure 2. 9 Tanglegram comparing the concatenated NifHDKENB tree with 16S rRNA phylogeny of the diazotrophs, both obtained using FastTree. Lines indicate the respective positions of the 963 bacteria in the 2 trees. See Supplementary Tree file for details of the horizontal gene transfer...61

Figure 2. 10 Proposed evolution of Diazotrophs. The gray tree represents the chronogram of prokaryotes while the green tree represents the proposed evolution of nitrogenase. The dashed green lines indicate horizontal gene transfer as supported by homology of NifHDKENB sequences. Black dots indicate origin of the respective phyla or groupings. The evolutionary timescale is based on Zhu et al., 2019 (Zhu et al., 2019). The Methanomada clade in Euryarchaeota includes orders Methanobacteriales and Methanococcales. Methanomicrobia I includes the orders Methanocellales and Methanomicrobiales. Methanomicrobia II includes Methanosarcinales. Clostridium I includes the family Peptococcaceae, Clostridium II includes Clostridia and Negativicutes.

Deltaproteobacteria I includes Myxococcales, and Deltaproteobacteria II includes other Deltaproteobacteria. 68

Figure 3. 1 The location of sampling sites at The Nature Conservancy (TNC) site at Sioux Prairie, South Dakota. Samples were collected from three parallel transects which were 5 meters apart. Source: google map 74

Figure 3. 2 Flow diagram illustrating isolation and processing of putative nitrogen fixing bacteria from Sioux Prairie soil samples. 77

Figure 3. 3 Colony characteristic of a typical isolate on R2A and NFA (left) and culturable count of putative diazotrophs incubated under aerobic and reduced oxygen condition at 28°C for 14 d (right)...83

Figure 3. 4 Diversity of potential diazotrophic bacteria by V1-V3 region of 16S rRNA sequences. Distribution of total isolates by phylum (a), distribution of phylum across three transect (b), and distribution of genera in the three different transects (c). 85

Figure 3. 5 Distribution of potential diazotrophs by incubation condition........................87

Figure 3. 6 Molecular phylogenetic analysis of partial 16S rRNA sequences obtained from potential diazotrophic isolates obtained by maximum likelihood in PhyML using aBayes analysis method. Branch tips are colored according to the genus classification and unclassified sequences not presented. The red stars on the outer ring indicates isolates which yielded nifH sequences on nifH PCR...88

Figure 3. 7 (a)Phylogenetic analysis of the nifH sequences obtained in this study constructed using PhyML. Size of black dots on the brach represent the aBayes branch

support. (b) Phylogeny using 926 nifH sequences obtained from previous study (chapter 2) as references. Tree constructed using RapidNJ with Kimura evolution model. The branches colored in red indicate sequences isolated in this study and nifH clusters are named according to (Zehr et al., 2003). 90

Figure 3. 8 $^{15}N_2$ assimilation of seven samples shows assimilation comparable to the negative control (E. coli). H. seropedicae was used as the positive control. 92

Figure 3. 9 Acetylene reduction assay of the selected isolates (left) and zoomed in (right) to show the ethylene peak. Presence of ethylene peak confirmed the diazotrophy.B16 (Paenibacillus), D8 (Bacillus), and Herb and E coli are positive and negative controls. ..93

Figure 4. 1 Colony characteristics of Herbaspirillium seropedicae and Streptomyces sp. strain 321G on NFA for 3 d and NFA supplemented with NH_3 for 24 hrs . Strain 321G in NFA+NH_3 (a) and NFA (b) and H. seropedicae in NFA+NH_3 (c) and NFA (d). 105

Figure 4. 2 Culture characteristics of strain 321G in NFM+NH_3 (a) and NFM (b) and H. seropedicae in NFM+NH_3 (c) and NFM (d). 106

Figure 4. 3 Streptomyces filaments under bright field microscope after staining with crystal violet. Streptomyces spores growing in NFM+NH_3 media after 24 h (a). Exponentially growing filaments in NFM+NH_3 transferred to NFM for 8 h (b), 24 h (c) and 7 d (d). 107

Figure 4. 4 Growth of Streptomyces sp. strain 321G estimated by absorbance by crystal violet after elution of the crystal violet bound to cell pellets in NFM (Top) and NFM+NH_3(bottom). 108

Figure 4. 5 Molecular phylogenetic tree of eight 16S rRNA sequences obtained from the assembled reads determined by Maximum-likelihood with 100 bootstrap replicates. The reference sequences are selected based on the closest alignment in SINA database.......110

Figure 4. 6 Functional classification of genes identified in the strain 321G. The genes were annotated and mapped by KEGG. 111

Figure 4. 7 The quality score and sequence count on each samples, obtained from fastQC on the raw fastq files. 113

Figure 4. 8 Taxonomy of raw fastq reads obtained from the nine samples given by Kaiju. Sample names are defined in Table 4.1. 114

Figure 4. 9 Sample distance (a, b) and principal component analysis (c) of the expression profile according to the different sampling condition..115

Figure 4. 10 Volcano plots showing differentially expressed genes by log2fold change and -log10(Adjusted p-value) between control Vs NFM(8 h) (a) and control Vs NFM (24 h) (b). Genes having adjusted p-value less than 0.05 and log2fold change greater than 1 are colored blue. 116

Figure 4. 11 Functional classification of differentially expresssed genes in eight-hour condtion (a) and 24-hour condition (b). Only the genes with more than 2 or less than -2 log2fold change are presented. 119

Figure 5. 1 The location of sampling site at the Nature Conservancy (TNC) site at Sioux Prairie. 4 sites 20 meters apart were selected for collecting the samples. Source google.map 130

Figure 5. 2 Diversity of prokaryotes based on the total RNA reads and the assembled transcripts. Diversity of the prokaryotic super-kingdoms (a) and the prokaryotic phyla excluding the unclassified reads (b)...135

Figure 5. 3 Diversity (a) and expression (b) of transcripts encoding structural nitrogenase components (nifHDKENB) 137

Figure 5. 4 Expression of the nitrification genes across the four sites. Oxidation of ammonia to nitrate occurs in two steps. Genes necessary for the conversion of ammonia to nitrite (AmoA and Hao) were detected but nitrite oxidation to nitrate was not detected. 138

Figure 5. 5 Phylogenetic association of the ammonia assimilating gene glnA(a) and the ammoina mineralizing gene deaminase (b). 139

Figure 5. 6 The measurement of expression of assimilation (glnA) and Mineralization (deaminase) across the four sites. The expression of glnA is found to be much higher than that of deaminase. 140

Figure 5. 7 Diversity of genes involved in multiple pathways of nitrate assimilation. ...142

Figure 5. 8 Expression of nitrate assimilation genes. 143

Figure 5. 9 Diversity (a) and expression (b) of marker genes involved in the denitrification process...144

Figure 5. 10 Nitrogen cycle showing log_2 expression of marker genes for each reaction as indicated by the thickness of the line (a). The expression level of N-cycling marker genese shown as log_{10} of average expresion of genes (b). 145

LIST OF TABLES

Table 3. 1 Primer pairs used for amplification of nifH genes in this study 80

Table 3. 2 Chemical analysis of NFM, different waters available in lab, arabinose and Noble agar used for solid NFM. 82

Table 4. 1 RNA quantities from nine samples passing the necessary requirements for sequencing.. 113

Table 4. 2 Differentially expressed pathways involved in nitrogen metabolism............ 123

Table 5. 1 GO terms used in this study to identify the transcripts involved in nitrogen cycling pathways.. 132

Table 5. 2 Quantification and quality of RNA extracted. 133

Table 5. 3 Quality of the sequencing reads.. 133

Table 5. 4 Trinity Assembly summary 134

ABSTRACT

FREE-LIVING DIAZOTROPHS AND THE NITROGEN CYCLE IN NATURAL GRASSLAND REVEALED BY CULTURE DEPENDENT AND INDEPENDENT APPROACHES

AMRIT KOIRALA

2021

Biological nitrogen fixation contributes to half of the global supply of nitrogen to the biosphere. It is carried out by a diverse group of prokaryotes called diazotrophs *via* the nitrogenase enzyme. Nitrogen fixation research is focused on the narrow group of symbiotic diazotrophs, and the vast majority of free-living diazotrophs which contribute significantly to fixed nitrogen are yet to be explored. The goal of this research was to access phylogeny of diazotrophs considering the most up-to-date genomic information and apply that knowledge to understand the diversity of free-living diazotrophs in a natural grassland ecosystem, both by culture dependent and independent methods.

Phylogeny was reconstructed using the concatenated sequences of six core proteins of nitrogenase (NifHDKENB) from 963 prokaryotic genomes. The diversity of free-living diazotrophs in grassland was explored by isolation of putative diazotrophs on a solid nitrogen free medium (NFM) and diazotrophy confirmed by *nifH* PCR, acetylene reduction assay and $^{15}N_2$ assimilation assay. *Streptomyces,* the most abundant bacteria, was further characterized by sequencing the genome of one prominent strain, and differential gene expression in nitrogen rich Vs nitrogen deficient medium. For culture-

independent study of nitrogen cycle activity, meta-transcriptomic sequencing of complete mRNA from a grassland soil sample was performed.

Phylogeny of *nif* genes from the complete genomes of cultured isolates revealed that diazotrophs are distributed across Actinobacteria, Aquificae, Bacteroidetes, Chlorobi, Chloroflexi, Cyanobacteria, Deferribacteres, Firmicutes, Fusobacteria, Nitrospira, Proteobacteria, PVC group, and Spirochaetes, as well as the Euryarchaeota, providing a curated database of *nif* genes. Culturing yielded 474 bacterial isolates which belonged to the phyla Actinobacteria, Proteobacteria, Firmicutes, and Bacteroidetes. However, only 81 (17%) of isolates yielded *nifH*, and the most dominant genus isolated on NFM, *Streptomyces* did not provide biochemical and genomic evidence of diazotrophy. The meta-transcriptomic study revealed nitrogen fixation and nitrification are the least and nitrate reduction is the most expressed pathway among various nitrogen cycling pathways.

In conclusion, although the culture-based approach showed diverse free-living nitrogen fixing bacteria, diazotrophy should always be confirmed by biochemical and genetic evidence, and limitations to culture independent study due to primer bias in *nifH* PCR can be overcome by meta-transcriptomic study.

Chapter 1

Literature Review

1.1 Introduction

Carbon, hydrogen, oxygen, and nitrogen are the four most abundant elements in a living cell. The cell mass is composed 70% or more of water and 80 to 90% of the dry weight is composed of macromolecules: carbohydrate, lipids, proteins, and nucleic acid (Cooper 2000). Of these macromolecules, nitrogen is a principal component of proteins and nucleic acids. Oxygen, carbon, and hydrogen can be easily cycled between the organic, biosphere and inorganic phase because of low activation energy necessary to make the conversions and the abundance of photosynthesis which utilizes photons absorbed from the sun to fix the carbon. However, nitrogen is not readily available to living organisms. Dinitrogen (N_2) which makes 78.1 % of the atmosphere is the major source of nitrogen for all living forms, but the triple bond between the two nitrogen atoms requires a bond dissociation energy of 226 kcal/mol, and is one of the strongest bonds found in nature (MacKay and Fryzuk 2004a). The process of conversion of atmospheric dinitrogen to reduced forms like ammonia (NH_3) which can be readily assimilated by living organism is called nitrogen fixation (Belnap 2002). It is carried out by very limited types of prokaryotes called diazotrophs, and by artificial methods like the Haber-Bosch process. As nitrogen is a major constituent of nucleic acids, amino acids, porphyrin (for example chlorophyll), and vitamins, and because nitrogen fixation is a very costly process, it is one of the important nutrients, limiting yields in terrestrial and aquatic ecosystems (Zehr et al. 2003).

Although a plethora of nitrogen is available in inert form in the atmosphere, the amount of nitrogen available for organisms is tightly regulated by nature by biological nitrogen fixation (Kim and Rees 1994). In contrast to chemical fixation, biological nitrogen fixation is regulated by a feedback loop mechanism which prevents excess nitrogen in the environment. Introduction of nonbiological nitrogen fixation through the Haber-Bosch process in the early 1900's brought the dawn of the "green revolution" which provided humankind with a substantial food supply. However, this bypass of natural process of nitrogen fixation has several disadvantages and environmental impacts (Klipp et al. 2005). Abundant supply of nitrogen fertilizers limits the reuse of animal manures which decreases the soil organic matter, degrading the physio-chemical quality like soil tilth, water-holding capacity, and the microbial population of soil. In addition to this, excess fertilizer is usually washed to water sources like rivers, streams, ponds, lakes, and coastal waters, causing eutrophication and loss of aquatic lives. Contamination of drinking water with these fertilizers causes several health hazards and oxidation of this excess ammonia to nitrate causes acidification of soil and environmental consequences like acid rain (Savcı 2012).

The International Nitrogen Management System (INMS) suggested the key to control nitrogen pollution is the rational use of chemical fertilizer and increasing the use of biological nitrogen fixation (Sutton et al. 2013). Since the discovery of nitrogen fixation in legume root nodules by Hellriegel and Wilfarth (Hellriegel and Wilfarth 1888), and isolation of diazotrophs from root nodules by Beijerinck (Beijerinck 1888) in the year 1888, several break throughs have marked progress in understanding of this natural mechanism that nature has devised to meet the nitrogen demand of the biosphere

(Ribbe 2011). However, the current knowledge about biological nitrogen fixation is still far from complete and several facets of this mechanism still needs to be deciphered (Hu and Ribbe 2014).

1.1.1 Global nitrogen cycle

Atmospheric nitrogen is the largest reservoir of nitrogen on the earth ($\sim 4 \times 10^9$ Tg N). However, it needs to be fixed before most organisms can assimilate it. The geological N reservoir in earth's crust as ammonium in silicate minerals, deep ocean sediments and sedimentary rocks ($\sim 2 \times 10^9$ Tg N) is in "fixed" form which is physically inaccessible to living organisms. Therefore, it contributes very little, if any nitrogen to the biosphere. The third most abundant source of nitrogen is ocean water which is estimated to contain $\sim 6.6 \times 10^5$ Tg N in the bioavailable form, most of which is in nitrate form. It is followed by biosphere ($\sim 1.6 \times 10^4$ Tg N), and soil/coastal sediments ($\sim 1 \times 10^4$ Tg N), both of which are in bioavailable form (Zhang, Ward, and Sigman 2020).

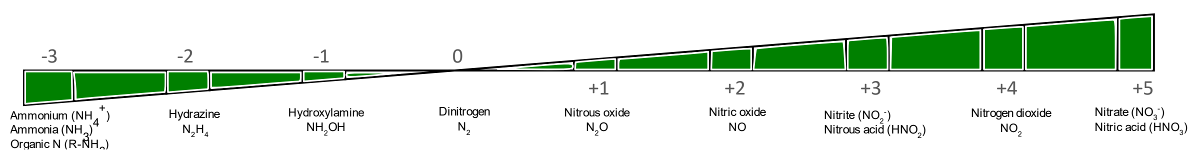


Figure 1. 1 Oxidation states of nitrogen in various stages in the global nitrogen cycle. Dinitrogen and Nitrous oxide with oxidation state zero and 1 respectively are non-fixed forms and not accessible to the majority of living organisms. The most reduced ammonium (-3) and the most oxidized nitrate (+5) are most easily assimilated by most of the organims and other radicals are formed in various stages in nitrogen cycle (Zhang et al. 2020).

The series of processes by which nitrogen is cycled between various oxidation states, among the above reservoirs by oxidization and reduction reactions is called as

global nitrogen cycle (Figure 1.1). Unlike any other nutrient cycles, the nitrogen cycle is unique because microorganisms play a vital role in both formation of bioavailable fixed nitrogen and removal of fixed nitrogen back to the atmosphere. Hence, the global nitrogen cycle is also frequently referred to as the microbial nitrogen cycle.

The nitrogen cycle can be described in three major processes (Figure 1.2).

a. Input process

Fixation of atmospheric dinitrogen into the reduced form ammonia/ammonium ion constitutes the input process and can occur by several natural and artificial methods.

i. Biological Nitrogen fixation: Nitrogen fixation by diazotrophs constitutes more than half of the global nitrogen input in the post-industrial age. It is estimated to contribute 128 Tg N yr^{-1} to the terrestrial ecosystem and 140 Tg N $yr^{-1}$to the marine ecosystem (Fowler et al. 2013).

ii. Haber Bosch process: It is one of the most exploited pathways in the modern agricultural world. It involves reaction of N_2 with H_2 in presence of ferrite catalyst at very high temperature (450ºC) and pressure (300 bar) to form ammonia which is used in commercial fertilizers. It is estimated to contribute 120 Tg N yr^{-1} to the terrestrial ecosystem (Fowler et al. 2015).

iii. Lightning: High temperature conditions associated with lightening also result in fixation of atmospheric nitrogen by forming reactive oxygen

radicals. It is estimated to contribute 5±3 Tg N yr^{-1} of nitrogen (Vitousek et al. 2013).

iv. Other: In addition to the above processes several natural processes like volcanic eruptions (0.9 Tg N yr^{-1}) (Paulot et al. 2015), hydrothermal vents, rock weathering (~ 10 to 30 Tg N yr^{-1}) (Houlton, Morford, and Dahlgren 2018) and human activities like fossil fuel combustion (~ 40 Tg N yr^{-1}) (Fowler et al. 2015) also add small quantities of fixed nitrogen to earth's surface.

b. Internal cycling

Once fixed, nitrogen is assimilated, mineralized, and reassimilated by living organisms which constitute the internal cycling of nitrogen. During this phase there is no net input from, or release of fixed nitrogen to atmospheric dinitrogen. It consists of the following processes as highlighted in Figure 1.2.

i. Nitrogen Assimilation: Ammonia is the immediate source of nitrogen for plants, phytoplankton, fungi and bacteria which incorporate it into amino acids by enzymes glutamine synthetase (glnA) and glutamate synthase (GLT1) (Eelen et al. 2018).

ii. Assimilatory Nitrate Reduction to Ammonia (ANRA): Although NH_3 is the preferred source of nitrogen, many organisms can utilize nitrate as nitrogen source by converting it to ammonium in a two-step reduction reaction. Nitrate is first reduced to nitrite by nitrate reductase and nitrite to

ammonia by nitrite reductase. Although many plants, fungi and algae are studied for NO_3^- assimilation, very few bacteria have been studied and significant difference has been observed in bacterial NO_3^--assimilation systems (Bothe, Ferguson, and Newton 2006). In photosynthetic bacteria like cyanobacteria both nitrate reductase (NarB) and nitrite reductase (NirA) are functionally linked to photosynthesis and use reduced ferredoxin (Fd) or flavodoxin (Fld) from photosynthesis as electron donor, whereas heterotrophic bacteria like *Pseudomonas, Bacillus, Azotobacter* use NADH dependent nitrate reductase (NR) and nitrite reductase (NIT-6). Also, in *Bacillus subtilis*, a gene cluster has been identified that codes for the catalytic subunit (NasA) and electron transfer subunit (NasB), which are deemed essential for nitrate assimilation in *B. subtilis* (Ogawa et al. 1995).

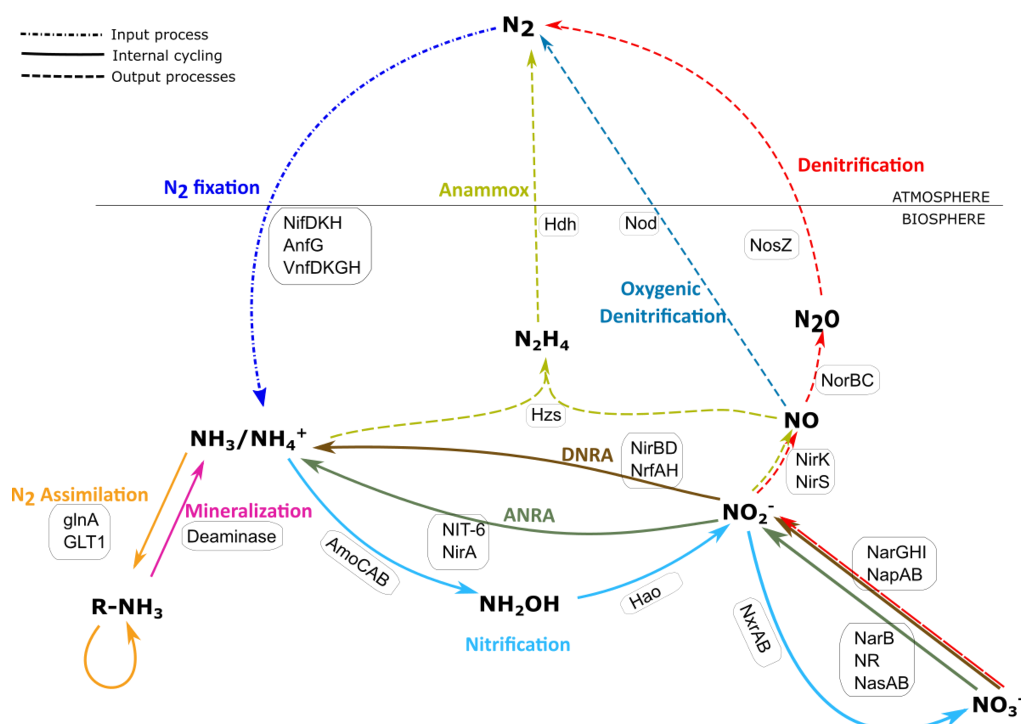


Figure 1. 2 Microbial nitrogen cycle highlighting the input process, internal cycling and output process and the microbial enzymes involved in each step. Adapted from Zhang et al. 2020 and the KEGG nitrogen metabolism map (Kanehisa and Goto 2000; Zhang et al. 2020).

iii. Mineralization: It is the process by which nitrogen assimilated by living organisms in organic form is returned to inorganic ammonium. It is accomplished by excretion of urea by higher organisms and degradation of dead and decaying living matter. Organic macromolecules are first broken down to dissolved organic nitrogen and followed by deamination of amino acids and nucleic acids.

iv. Nitrification: It is the process by which ammonia is oxidized to nitrate in presence of oxygen. Classically, it is carried out by a group of chemolithoautotrophs which require ammonia or nitrite as source of electrons, energy, and biomass nitrogen. Ammonia oxidizers like *Nitrosomonas* and *Nitrosopumilus* oxidize ammonia to nitrite in two step

reactions where ammonia monooxygenase (AmoCAB) converts ammonia to hydroxylamine and hydroxylamine dehydrogenase (Hao) catalyzes conversion to nitrite. Nitrite oxidizing bacteria like *Nitrobacter* and *Nitrospina* oxidize nitrite to nitrate by the enzyme nitrite oxidoreductase (NxrAB) in presence of oxygen. A recently identified group of proteobacteria called Comammox harbor all the enzymes necessary for complete oxidation of ammonia to nitrate (van Kessel et al. 2015).

v. <u>Dissimilatory Nitrate Reduction to Ammonia (DNRA)</u>: It is reduction of nitrate to ammonia by chemoheterotrophic bacteria where nitrate is first reduced to nitrite and then nitrite is used as electron donor for anaerobic respiration for oxidation of organic matter. Enzymes used for reduction of nitrate to nitrite are common with denitrification, however DNRA is unique because it cycles fixed nitrogen back to ammonia in contrast to denitrification where nitrate is lost from the biosphere as dinitrogen. Nitrate reduction in DNRA is carried out by two types of nitrate reductase. NapAB is the membrane-bound, periplasm facing enzyme, and NarGHI is membrane-bound, cytoplasm-facing nitrate reductase. The next step is reduction of nitrite to ammonia which is carried out by either NADH dependent nitrate reductase (NirBD) or cytochrome dependent nitrate reductase (NrfAH).

c. Output process

This is the process by which bio-available nitrogen is released back to the atmosphere as dinitrogen. It consists of the following steps:

i. Denitrification: It is the primary route by which fixed nitrogen is converted back to N_2. It is estimated to remove 100 Tg N yr^{-1} from the terrestrial ecosystem and 100-380 Tg N yr^{-1} from the marine ecosystem (Fowler et al. 2013). It involves reduction of nitrate to nitrite by the same enzymes involved in DNRA, followed by reduction of nitrite>nitric oxide (NO) >nitrous oxide (N_2O) > dinitrogen. Reduction of nitrite to nitric oxide is carried out by one to two distinct enzymes; NirK contains copper in its active site and NirS is a cytochrome-containing Fe protein. Nitric oxide reduction to nitrous oxide is carried out by nitric oxide reductase (NorBC), and nitrous oxide is reduced to dinitrogen by the complex copper enzyme NosZ.

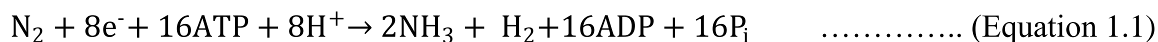
ii. Anammox: It is the oxidation of ammonium using nitrite as electron acceptor and bacteria belonging to the phylum *Planctomycetes* carry out this transformation using the enzyme hydrazine synthase (Hzs) to form the intermediates N_2H_2, and hydroxylamine dehydrogenase (Hdh) to form dinitrogen from N_2H_2.

iii. Oxygenic denitrification: It is a new route in release of nitrate back to dinitrogen and is marked by direct disproportionation of NO into N_2 and

O_2. It is catalyzed by the putative NO dismutase (Nod) (B. Zhu et al. 2019).

1.1.2 Biological Nitrogen Fixation (BNF)

Although the most abundant form of nitrogen, dinitrogen in atmosphere, is unavailable to living organisms because of strong bond energy (MacKay and Fryzuk 2004b). Nature has devised BNF which uses a large, structurally complex, iron-rich metalloenzyme called nitrogenase to overcome the activation energy required for the conversion to ammonia (Postgate, Stewart, and Rosswall 1982). From the origin of nitrogenase enzyme approximately two billion years ago (Boyd and Peters 2013) to the pre-industrial age, BNF was the only major contributor of fixed nitrogen in the global nitrogen cycle. In the post-industrial era BNF still contributes half of the terrestrial fixed nitrogen supply, and in the marine ecosystem it is still the sole major contributor of fixed nitrogen (Zhang et al. 2020). Compared to the artificial Haber-Bosch process which needs very high temperature (~450ºC) and pressure (>200 atm) and utilizes ~1 % of global fossil fuel (Smith 2002), BNF is carried out by diazotrophs at ambient temperature and pressure with the energy derived from biological sources (ATPs) (Ribbe 2011). The overall reaction of nitrogenase catalyzed nitrogen fixation according to kinetic scheme (Howard and Rees 2006) is given by equation 1.1.



$$N_2 + 8e^- + 16ATP + 8H^+ \rightarrow 2NH_3 + H_2 + 16ADP + 16P_i \quad \text{............. (Equation 1.1)}$$

For each molecule of dinitrogen reduced to two molecules of ammonia, there is obligatory production of one molecule of hydrogen (Simpson and Burris 1984). Of the 16

ATPs per N_2 reduced, four ATPs are used for the generation of one hydrogen molecule (Hoffman et al. 2014).

1.1.3 Nitrogenase complex

The only known and extensively studied enzyme system that catalyzes BNF is the nitrogenase enzyme complex (Burgess and Lowe 1996). It is a multiprotein complex enzyme which accounts for 10% of the cell protein in diazotrophs (Dingler et al. 1988). Nitrogenase is found in a wide range of organisms with significant sequence diversity, however, all of them have the conserved structure of a two-component system (Bulen and LeComte 1966). It has the ability to reduce many other small molecules like acetylene (C_2H_2), hydrogen cyanide (HCN), and carbon monoxide (CO) (Hoffman et al. 2014) and is irreversibly inactivated by O_2 by binding to the Fe protein cluster (Gallon 1981). A second type of nitrogenase which is insensitive to O_2 and not inhibited by presence of CO and H_2 and does not catalyze the reduction of acetylene to ethylene was reported in *Streptomyces thermoautotrophicus* (Gadkari, Mörsdorf, and Meyer 1992). However, recent reports focused on characterization of this nitrogenase refute the existence of such an enzyme (MacKellar et al. 2016).

1.1.3.1 Structure and types

Nitrogenase is a two-component system (Figure 1.3) composed of Component I (dinitrogenase or MoFe protein) and Component II (dinitrogenase reductase or Fe protein). The MoFe protein contains two metal clusters: the iron-molybdenum cofactor (FeMo-co), and P-cluster. The Fe protein contains one Fe-S cofactor. There are three related forms of nitrogenase based on the metal in FeMo-co. The most abundant and

well-studied nitrogenase is molybdenum based nitrogenase for which crystallographic structures were solved as early as 1996 (Howard and Rees 1996), and has Mo in FeMo-co. The alternative forms of nitrogenase are V-nitrogenase (Mo in FeMo-co replaced by V) and Fe-nitrogenase (has FeFe-co instead of FeMo-co) (Eady 1996).

The Mo-nitrogenase complex is composed of a central MoFe protein (~240 kDa) which is a hetero tetramer of two α chains and two β chains. The Fe protein is a homodimer with molecular weight 60-70 kDa and bound to each αβ half of MoFe protein. The arrangement of two Fe proteins around one MoFe cluster is frequently referred as 2:1 Fe protein-MoFe protein complex (Jang, Seefeldt, and Peters 2000), observed when crystallized in the presence of excess MgADP. Fe protein serves to couple ATP hydrolysis with interprotein transfer of electrons. The F cluster in Fe protein is positioned at the junction of two identical Fe protein subunits and MoFe protein. It transfers an electron to the P-cluster [8Fe-7S] located on the surface of the MoFe protein at the interface of α and β subunits within each αβ half, which in turn transfer it to the MoFe-co buried in the active site 10 Å beneath the surface of each α subunit (Burén et al. 2020).

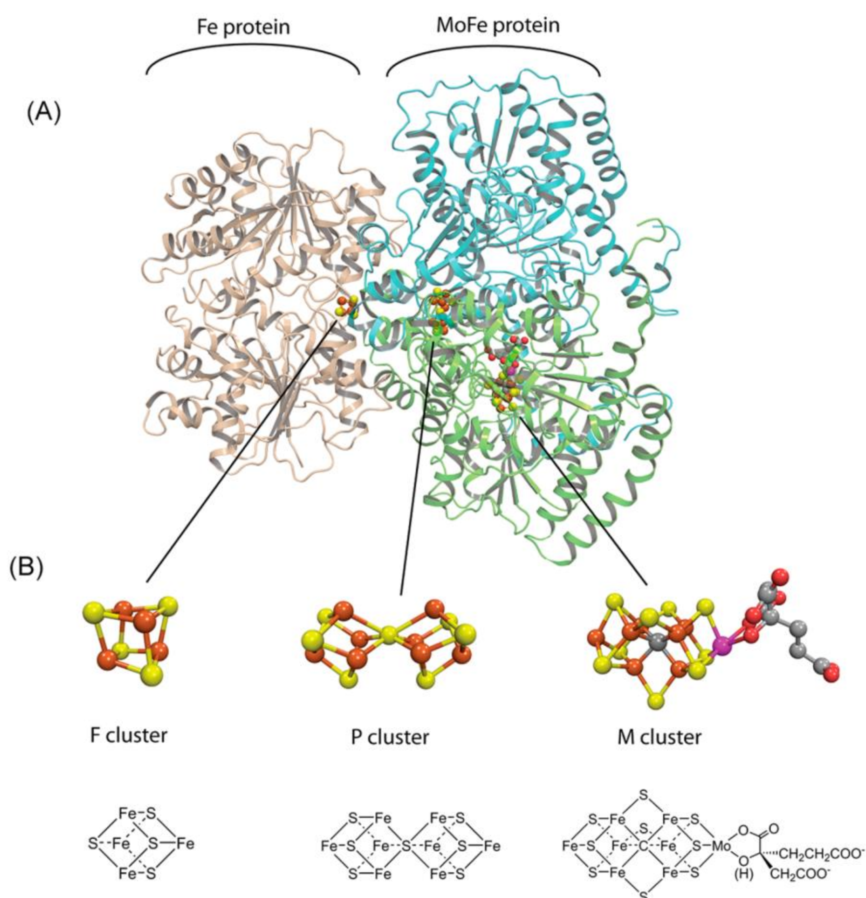
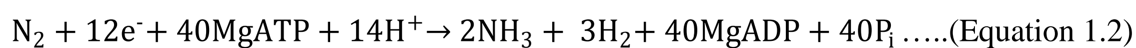


Figure 1. 3 Structure of one catalytic half of Molybdenum nitrogenase. (A) Component II (Fe protein homodimer) shown in tan, and Component I (dinitrogenase) shown in green (α subunit) and cyan (β subunit). (B) Fe protein has one Fe-S cluster and MoFe protein has P-cluster and a MoFe co (M-cluster). The M cluster has an inorganic [7Fe-9S-C-Mo] cluster and a molecule of R-homocitrate (Hoffman et al. 2014).

The amino acid sequence around the cofactor binding sites is highly conserved in both MoFe protein and Fe protein (Newton 2015). In the *Azotobacter vinelandii* MoFe protein, FeMo-co is ligated to the α-Cys^{275} residue to its terminal Fe, and by an α-His^{442} residue to the Mo atom, and most of the amino acids around the active site are hydrophilic. Similarly, the conformation of the P-cluster depends on the oxidation state. The oxidized P-cluster in *A. vinelandii* is covalently linked to six Cys residues (Cys^{62}, Cys^{88}, Cys^{154} of the α subunit and Cys^{70}, Cys^{95}, Cys^{153} of the β subunit). The F cluster in

Fe protein is coordinated by two Cys residues from each subunit (Cys^{97} and Cys^{132} in *A. vinelandii*) (Burén et al. 2020).

The vanadium nitrogenase has very similar sequence structure to the Mo-nitrogenase and has the same two-component structure but with an additional δ-subunit which is homologous to the NifY protein in Mo-nitrogenase. In Mo-nitrogenase, NifY is detached in the maturation stage, but it remains as VnfG in the mature V-nitrogenase. Fe protein in V-nitrogenase (molecular weight ~64 kDa) is a homodimer having 91% sequence homology with the Fe protein of Mo-nitrogenase (Hu, Lee, and Ribbe 2012). The Fe protein from V-nitrogenase also has a conserved Gly-X-Gly-X-X-Gly sequence at the MgATP binding site, and a conserved Cys ligand at the cofactor binding site (Rehder 2000). The VFe protein from *A. chrochococcum* is ~250 kDa, and has an $\alpha_2\beta_2\delta_2$ heterohexameric structure (Eady 1996), but recently VFe protein isolated from *A. vinelandii* was found to be a hetero-octameric $\alpha_2\beta_2\delta_4$ with molecular weight ~270 kDa (Lee, Hu, and Ribbe 2009). The α and β subunits of the VFe protein show ~33% and ~32% sequence homology with respective proteins in the MoFe protein (Hu et al. 2012). The overall reaction of N_2 fixation carried out by V-nitrogenase is given by equation 1.2 (Eady 2003).



$$N_2 + 12e^- + 40MgATP + 14H^+ \rightarrow 2NH_3 + 3H_2 + 40MgADP + 40P_i \quad \text{.....(Equation 1.2)}$$

The reduction of N_2 to NH_3 by V-nitrogenase is less efficient where only 50% of the electron flux is used to reduce nitrogen, compared to 75% in Mo-nitrogenase (Schneider et al. 1997). V-nitrogenase is also able to reduce other small molecules like C_2H_2, CO and CO_2 but with much lower efficiency (Yang, Dean, and Seefeldt 2011).

The iron-only nitrogenase has FeFe-co at the active site and is the least efficient nitrogenase which is expressed only in absence of both Mo-nitrogenase and V-nitrogenase (Hinnemann and Nørskov 2004). Only 30% of the electron flux is used towards reduction of N_2 in Fe-nitrogenase (Schneider et al. 1997). Because of extreme sensitivity to O_2, Fe-nitrogenase is very unstable and has been isolated from *A. vinelandii* (Eady 2003), *Rhodospirillum rubrum* (Lehman and Roberts 1991), and *Rhodobacter capsulatus* (Schneider et al. 1991). The Fe protein of Fe-molybdenum has ~60% homology with Fe protein of Mo-molybdenum (Eady 1996). The structure of the FeFe protein of Fe-nitrogenase is very similar to the VFe protein of V-nitrogenase and is found to be a hetero-hexamer of $\alpha_2\beta_2\delta_2$ (Eady 1996).

1.1.3.2 Mechanism of reduction of N_2 by Mo-nitrogenase

The reduction of N_2 to yield two molecules of ammonia is thermodynamically favorable but a large quantity of MgATP is required for the functioning of nitrogenase for kinetic reasons. MgATP drives one way flow of electrons from the nitrogenase reductase (Fe protein) to the active site in the MoFe protein, subsequently accumulating multiple electrons at the MoFe-co and reducing N_2 to NH_3 (Lanzilotta and Seefeldt 1997). This kinetic model was extensively studied and developed by Lowe and Thorneley and is known as the Lowe-Thorneley (LT) kinetic model for nitrogenase function (Burgess and Lowe 1996; Thorneley and Lowe 1996). According to this model reduction of N_2 to NH_3 occurs in the following stages:

i. Fe protein-MoFe protein complex formation is an essential step in functioning of nitrogenase and occurs for only 1 second during normal

substrate reduction. Hydrolysis of two MgATP and transfer of one electron from Fe protein to MoFe protein occurs during this complex formation. The hydrolysis of two MgATP molecules causes movement of amino acids in positions called switches I and II in the Fe protein which changes the distance between the F-cluster in Fe protein and the P-cluster in the MoFe protein by 5 Å, inducing the electron transfer and essentially reducing the MoFe protein (Tezcan et al. 2005).

ii. The Fe protein cycle is the next step where repeated association and dissociation of Fe protein with MoFe protein causes accumulation of electrons in the active site (MoFe-co) of the MoFe protein. Since eight electrons are required for reduction of N_2 to NH_3 and only one electron is transferred per association-dissociation event, multiple association and dissociation of Fe protein and MoFe protein occur (Figure 1.4).

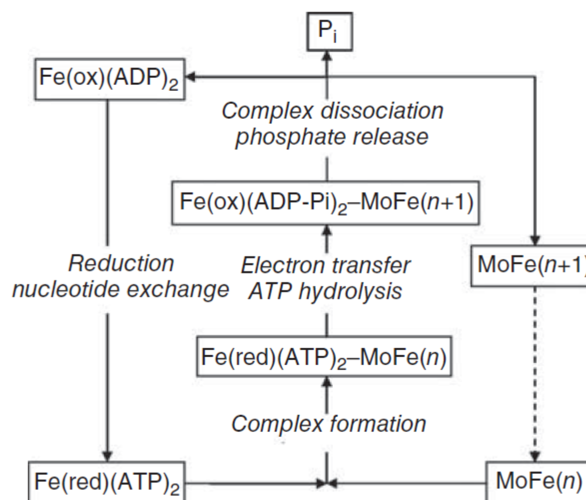


Figure 1. 4 Fe protein cycle showing the multistep reduction of MoFe protein by Fe protein. The reduced Fe protein (Fe(red)(ATP)$_2$ forms the complex with MoFe protein (MoFe(n): where n= number of electrons in the active site MoFe protein). During the dissociation of the complex Fe protein is oxidized (Fe(ox)ADP$_2$) and MoFe protein gains one more electron (MoFe(n+1)). The oxidized Fe protein is cycled again by reduction by nucleotide exchange and the reduced MoFe(n+1) can either bind to substrate or keep on gaining electrons by forming the complex with recycled Fe protein. From (Newton 2015).

iii. MoFe protein cycle involves progressive reduction of MoFe protein by up to eight electrons during eight Fe protein cycles (Figure 1.5). For simplicity MoFe protein is represented as E_n where n= number of electrons in the MoFe protein (including both P and M clusters). E_0 is the oxidized form of MoFe protein with zero electrons, and in each Fe protein cycle it gains one electron and one proton. For N_2 to be bound to MoFe protein it must be in E_3 or E_4 states which is accompanied by the release of one equivalent of H_2.

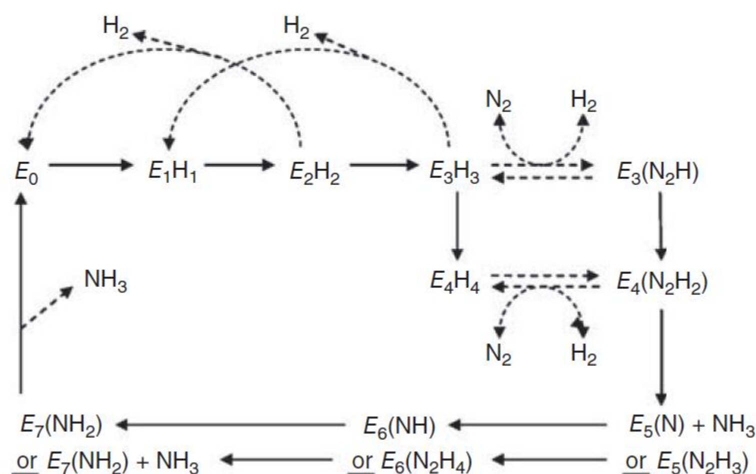


Figure 1. 5 MoFe protein cycle as depicited by the Lowe and Thorneley model. MoFe protein is cycled between eight oxidation states (E_0 to E_7). Different substrates of nitrogenase bind reversibly to differnet MoFe protein redox states. For example N_2 binds to the E_3 or E_4 state and C_2H_2 binds to either the E_1 or E_2 state. From (Hoffman et al. 2014)

1.1.3.3 Genomic organization of nitrogenase genes

Genes coding for Mo-nitrogenase, V-nitrogenase, and Fe-only nitrogenase are termed *nif, vnf,* and *anf* genes respectively (Eady 1996). The structural components of the Mo-nitrogenase: Fe protein (NifH) is coded by *nifH*, and components of the MoFe protein, α-chain (NifD) and β–chain (NifK) is coded by the *nifD* and *nifK* genes respectively. Similarly, genes coding for structural components of V-nitrogenase are *vnfHDK* and for Fe-only nitrogenase are *anf*HDK. The additional component, δ-subunit present in the alternative forms of nitrogenase is coded by a *vnf/anfG* gene for the VFe protein and FeFe protein respectively. Genes coding for structural components of nitrogenase occur in close association with other accessory genes necessary for the assembly of nitrogenase and regulation of nitrogen fixation in *nif* gene clusters with some

variation among different organisms(Roberts et al. 1978; Rodríguez-Quiñones, Bosch, and Imperial 1993; Setubal et al. 2009). For example all the *nif* genes in *Klebsiella pneumoniae* occur on a single *nif* gene cluster containing twenty *nif genes (nifJHDKTYENXUSVWZMFLABQ)*, and in *A. vinelandii* they occur in two clusters named major (35 *nif* genes including the structural *nifHDK*) and minor (17 *nif* genes) *nif* cluster based on their size (Jacobson et al. 1989) (Figure 1.6). The accessory genes present in *nif* clusters vary according to the different physiological growth conditions. For example, *A. vinelandii* carries additional genes to protect nitrogenase from oxygen, compared to *K. pneumoniae* which fixes nitrogen only in strict anaerobic conditions.

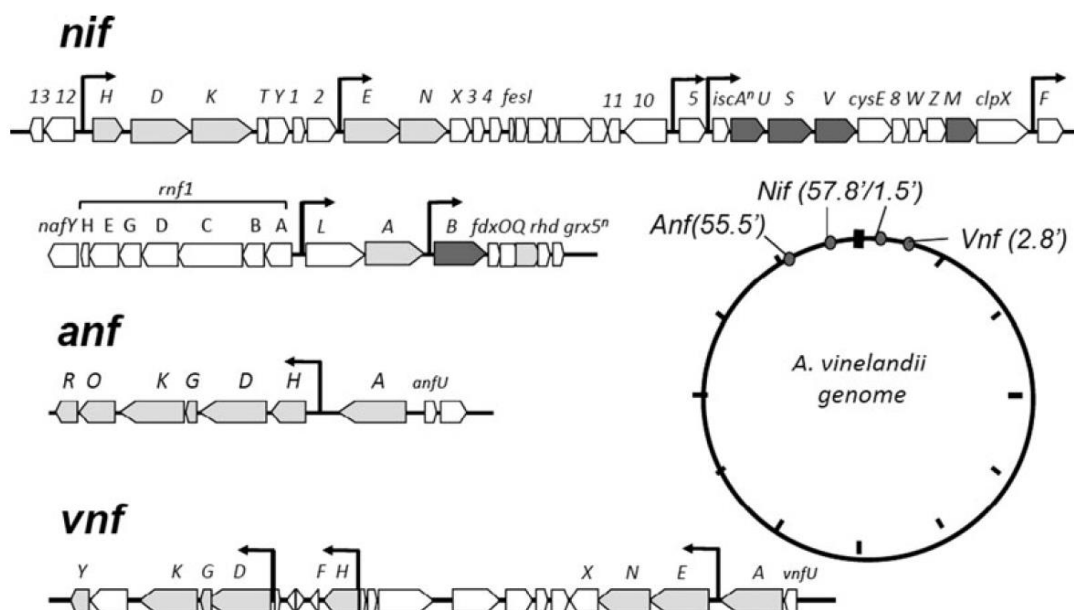


Figure 1. 6 Genomic arrangement of nitrogen fixing genes in A. vinelandii. The nif genes occur in two clusters and A. vinelandii has separate anf and vnf cluster as well. From (Ribbe 2011)

As MoFe-co is one of the most complex cofactors found in living organisms, the assembly and maturation of M, P and F-cluster in nitrogenase requires multiple accessory

proteins. At least 12 *nif* genes are found to be involved in synthesis, assembly and maturation of these cofactors (Rubio and Ludden 2005). Other accessory proteins like flavodoxin (NifF), pyruvate:flavodoxin oxidoreductase (NifJ) provide nitrogenase with electrons, and NifA and NifL are involved in the regulation of *nif* gene expression (Martinez-Argudo et al. 2004).

Organisms capable of expressing alternative forms of nitrogenase have separate cluster for respective nitrogenases (Addo and Dos Santos 2020). *A. vinelandii* has *anf* and *vnf* clusters in addition to the *nif* clusters (Figure 1.6). The *vnf* cluster contains genes for structural components of V-nitrogeanse (*vnfHDKG*), and incomplete assembly proteins (*vnfEN, vnfX, vnfY,* and *vnfU*). The *anf* cluster is even smaller with genes coding for structural genes (*anfHDGK*), a transcriptional regulator (*anfA*) and two genes that code for proteins of unknown function (*anfAU)* (Setubal et al. 2009).

The distribution of genes associated with nitrogen fixation has be extensively studied by Dos Santos et al. (Addo and Dos Santos 2020; Dos Santos et al. 2012). Based on a study of 2000 complete genomes available in public database they proposed presence of *nifHDKENB* as the minimum criteria for computational prediction of diazotrophy.

1.1.3.4 Regulation of BNF

Biological nitrogen fixation is a very energy intensive process and diazotrophs have to spend a significant amount of cellular resources to maintain the repertoire of gene products for the synthesis, assembly and functioning of nitrogenase (Postgate et al. 1982). In addition to this, organisms growing in aerobic or microaerophilic conditions have to

spend additional resource to maintain an intracellular anaerobic environment using highly active respiratory chains (Bertsova, Bogachev, and Skulachev 2001). Hence, BNF is precisely regulated by complex regulatory cascades at transcriptional and post-translational level (Dixon and Kahn 2004a).

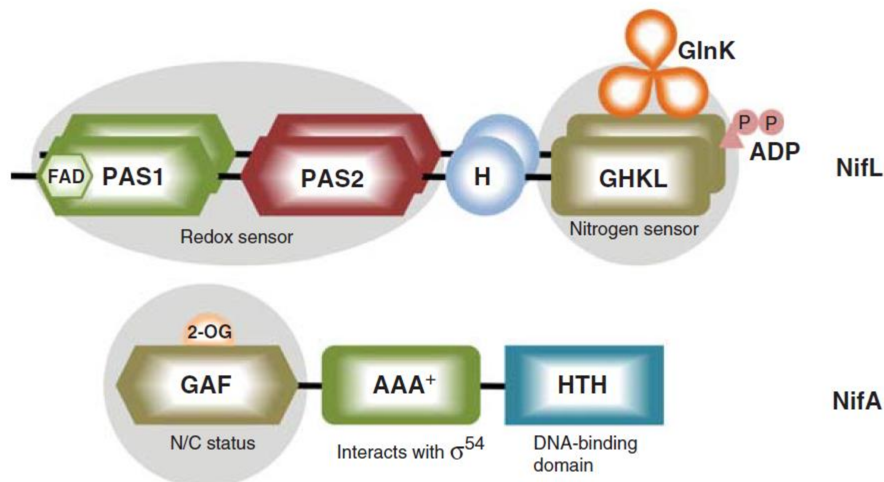


Figure 1. 7 The NifL and NifA system acts as regulator of nif gene expression. NifA is the activator of RNA polymerase σ^{54} factor and NifL is the anti-activator of NifA. Both of these proteins have multiple domains to receive the signal about redox state, C:N ratio and ATP level. From (Poza-Carrión, Echavarri-Erasun, and Rubio 2015a)

Regulation of nitrogen fixation has been extensively studied in Proteobacteria like *Azotobacter, Klebsiella,* and *Desulfovibrio*, and Cyanobacteria like *Trichodesmium* and Firmicutes like *Clostridium* (Zhang et al. 2020). In most of the diazotrophs, the NifA-NifL system is the major transcriptional regulator of nitrogenase. Enhancer-binding protein NifA (the master regulator of *nif* genes) is a transcriptional activator which binds to upstream activator sequences and facilitates the binding of RNA polymerase σ^{54} factor (Buck et al. 1986). NifL on the other hand is an anti-activator of NifA and together these two proteins regulate transcription of *nif* genes in response to the following major factors:

i. Intracellular redox potential: The N-terminal domain of NifL carries flavin adenine dinucleotide (FAD) cofactor which remains in oxidized state in presence of oxygen and reduced state in absence of oxygen. When FAD is oxidized it makes conformational changes in NifL which will result in binding of NifL with NifA, thereby inhibiting the activity of NifA and decreasing *nif* transcription.

ii. Cellular energy level: The C-terminal domain of NifL is an ATP-binding domain and can sense the cellular energy level.

iii. Carbon:Nitrogen balance: The C-terminal domain also interacts with GlnK which occurs in either non-uridylylated (high N:C ratio condition) or uridylylated (low N: C ratio condition). The non-uridylylated GlnK can bind to the C-terminal domain of NifL and inhibit the activity of NifA.

The N-terminus of NifA has a GAF domain which can bind with 2-oxoglutarate (2-OG) there, by preventing the binding of NifL. 2-OG acts as the direct measure of carbon available in the cell.

The ideal condition for nitrogen fixation is high adenylate energy change (AEC), high C:N ratio and low redox potential. AEC is calculated from ATP, ADP, and AMP from the Equation 1.3 (Upchurch and Mortenson 1980). During these condition NifL is unable to bind to NifA, so NifA can activate the RNA polymerase by binding to the activator sequence (Martinez-Argudo et al. 2004).

$$\mathrm{AEC} = \frac{[\mathrm{ATP}]+0.5[\mathrm{ADP}]}{[\mathrm{AMP}]+[\mathrm{ADP}]+[\mathrm{ATP}]} \quad \ldots\ldots\ldots\ldots \quad \text{(Equation 1.3)}$$

1.1.3.5 Phylogeny of nitrogenase

Nitrogenase is homologous to the enzyme protochlorophyllide reductase which is also a complex protein made up of three components Chl/BchL, Chl/BchN, and Chl/BchB. ChlLNB protochlorophyllide reductase is found in cyanobacteria, algae, and gymnosperms, and BchLNB is found in photosynthetic bacteria. Protochlorophyllide reductase plays a key role in light-independent chlorophyll biosynthesis. It catalyzes the reduction of ring D of protochlorophyllide (Pchlide) to form chlorophllide a (Chlide) (Fujita and Bauer 2000).

Figure 1. 8 Phylogeny of nitrogenase and homologous using NifHD protein. It shows five groups of which functional nitrogenase were found only in first three groups. Group IV contains uncharacterized Nif proteins and Group V contains protochlorophyllide reductase proteins.From (Raymond et al. 2004).

The structural component of nitrogenase reductase, Nif/Vnf/AnfH belongs to the NifH/frxC Pfam family (PF00142) along with ChlL(frxC) and BchL proteins. Similarly, the structural components of dinitrogenase Nif/Vnf/AnfDK belong to the Oxidoreductase family (PF00148) along with Chl/BclNB (Mistry et al. 2021). One of the earliest and the most followed phylogenies of nitrogenase is the one proposed by Raymond et al in 2004 (Raymond et al. 2004) which is based on NifHD proteins obtained from 110 publicly available genomes at that time. According to this classification nitrogenase and its homologs cluster into five distinct clusters (Figure 1.8).

Group I contain Mo-nitrogenase from most aerobic and facultative anaerobic cyanobacteria and proteobacteria. These are the most studied nitrogenase and each organism has developed spatial and/or temporal mechanisms for protection of its nitrogenase from oxygen. The phylogeny of nitrogenase follows 16S rRNA phylogeny with some mismatches, indicating the existence of horizontal gene transfer events (details in chapter 2).

Group II comprises of Mo-nitrogenase from strictly anaerobic bacteria like *Clostridium* and sulfate-reducing bacteria and archaea like *Methanosarcina.* All these organisms are frequently found in syntrophic ecosystems and in close association which should have facilitated horizontal gene transfer between these phylogenetically distinct taxa (Chien and Zinder 1996).

Group III consists of the alternative nitrogenases V-nitrogenase and Fe-only nitrogenase. Although Vnf/AnfDK are phylogenetically distinct from NifDK, Vnf/AnfH cannot be distinguished from NifH phylogenetically and most of the time the Vnf/AnfH

annotation is based on the position of these genes closer to a *vnf/anfDK* operon. The deeper branching of Group III suggests that alternative nitrogenases evolved earlier than Mo-nitrogenase. However phylogenetic analysis by Boyd and Peters, 2013 found that alternative nitrogenases are nested among other Mo-nitrogenases suggesting that alternative nitrogenases have evolved from Mo-nitrogenase (Boyd and Peters 2013).

Group IV is characterized by Nif homologs called Nif-like proteins (NflHDK) which are not characterized yet and are not involved in nitrogen fixation. Although NflH has some conserved residues at the Fe-S cluster and P-loop, NflD and NflK are highly variable and residues at the metal cofactor binding sites are also not conserved. A recent report has suggested evidence of nitrogen fixation by the Group IV nitrogenase in *Endomicrobium proavitum* (Zheng et al. 2016).

Group V consists of protochlorophyllide reductase which are homologs but have no role in nitrogen fixation.

In addition to this phylogeny based on NifHD, NifH phylogeny on its own is also frequently used for the study of diazotrophs (Gaby and Buckley 2012a; Zehr et al. 2003). The NifH phylogeny proposed by Zehr is in close agreement with Chien and Zinder (Chien and Zinder 1996), and is actively maintained and updated at https://www.jzehrlab.com/nifh (Zehr et al. 2003) (Figure 1.9).

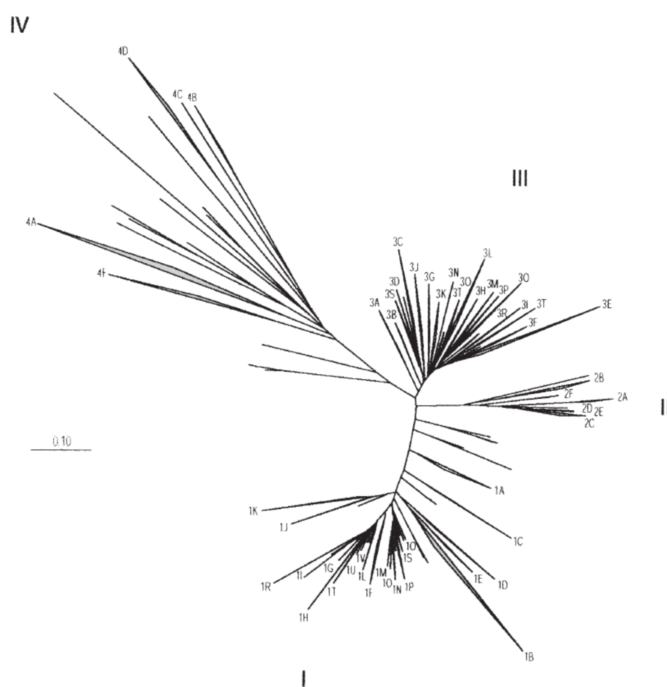


Figure 1. 9 NifH phylogeny given by Zehr et al. It has four clusters which show loose association with 16S rRNA phylogeny. From (Zehr et al. 2003).

According to this classification scheme, Cluster I consist of NifH and VnfH from aerobic and facultative anaerobic proteobacteria and cyanobacteria. Cluster II consists of VnfH and NifH from some archaea. Cluster III consists of NifH from strict anaerobes and Cluster IV includes all NifH homologs like NflH and Bch/ChlL.

1.1.4 Diversity of diazotrophs

Diazotrophs are distributed sporadically among prokaryotes in Bacteria and Archaea (Young 2005). So far around 100 genera from Proteobacteria, Cyanobacteria, Chlorobi, Spirochaetes, Firmicutes, Actinobacteria, Bacteroidetes, Nitrospirae, and Verrucomicrobia are shown to fix nitrogen and some species of Fusobacteria, Deferribacteres, and Fibrobacteres are shown to possess *nif* genes (Rosenberg et al. 2013) . However, diazotrophic species are distributed in patches among non-nitrogen fixing species. In some taxa, there is only one genus fixing nitrogen, for example *Frankia* in the Actinobacteria (Normand et al. 1996), *Paenibacillus* in the aerobic Firmicutes (Achouak, Normand, and Heulin 1999) and *Gluconacetobacter diazotrophicus* and a few others in a larger group comprising *Acetobacter, Gluconoacetobacter,* and *Gluconobacter* (Fuentes-Ramírez et al. 2001). Based on their ecological distribution, diazotrophs can be classified as one of the following:

i. Symbiotic diazotrophs form a permanent association with the host and are characterized by morphological and physiological modification (Vance 2001). They depend on the host for energy and protection from oxygen. The classical example of symbiotic nitrogen fixation is the formation and colonization of root nodules in leguminous plants by a diverse group of bacteria including *Rhizobium, Mesorhizobium, Bradyrhizobium, Burkholderia, Ensifer, Sinorhizobium* etc from Alpha and Betaproteobacteria termed rhizobia (Rascio and La Rocca 2008). The host plant provides carbon and protection from oxygen and the bacterial

counterpart fix nitrogen. Nitrogen fixing root nodules are also observed in actinorhizal plants where Actinobacteria (*Frankia)* occur in symbiosis. Similarly, several cyanobacteria form bipartite and tripartite symbiosis with algae, fungi, ferns, and mosses. For example, *Nostoc-Gunnera* (Cyanobacteria and plant) and *Richelia-Rhizosolenia, and Hemiaulus* (Cyanobacteria and diatom) (Wouters et al. 2009). A unique marine unicellular symbiosis between the uncultured unicellular cyanobacterium UCYN-A and a haptophyte picoplankton alga has been reported in oligotrophic oceans (Zehr et al. 2016).

A different form of symbiosis is observed in the insect gut where bacterial species like *Treponema* (Graber, Leadbetter, and Breznak 2004), *Diplosphaera* (Wertz et al. 2012), and *Candidatus* Azobacteroides pseudotrichonymphae (Inoue et al. 2015) occur. Another eukaryote, the bivalve *Codakia orbicularis* is found to live in symbiotic association with a sulfur-oxidizing bacterium, *Candidatus* Thiodiazotropha endolucinida in its gills (König et al. 2016).

ii. Free-living diazotrophs which can fix nitrogen on their own in ambient growing condition. This includes most of the marine diazotrophs (Zehr and Capone 2020), terrestrial cyanobacteria (Esteves-Ferreira et al. 2017), methanogenic Archaea (Adam et al. 2017), and Firmicutes (Achouak et al. 1999). Since nitrogenase is irreversibly inactivated by O_2, and most of these diazotrophs grow in aerobic condition, free-living diazotrophs need

to spend additional cellular resources to decrease redox potential around nitrogenase. Hence, free-living diazotrophs are less efficient and fix about one-tenth of the total atmospheric N_2 fixed by symbiotic diazotrophs (Unkovich and Baldock 2008). However, given the omnipresence of these bacteria in a wide array of habitats, nitrogen fixation by free-living diazotrophs has been predicted to contribute a significant amount of fixed nitrogen towards biological nitrogen fixation (Vitousek et al. 2013). Some diazotrophs belonging to genera *Paenibacillus, Herbaspirillium, Klebsiella,* and *Azospirillum* occur in close association with plants in the rhizosphere, and are termed associative diazotrophs and depend on plant exudates for carbon source (Van Dommelen and Vanderleyden 2007). Free-living diazotrophs and associative diazotrophs are together termed as asymbiotic nitrogen fixers.

1.1.5 Approaches to study diversity of diazotrophs

1.1.5.1 Culture dependent methods

Bacteria in culture are the primary requirements for the study of physiological and metabolic traits of any organism, and cultures are considered the gold standard in diagnosis of various pathogens (Abayasekara et al. 2017). Although 99% of the soil bacteria were estimated to be unculturable by standard techniques, recent understanding of bacterial physiology, advances in culture techniques, and availability of new media ingredients have made it possible to grow various bacteria previously thought

unculturable in the laboratory (Cardenas and Tiedje 2008). Thus culture-independent study of nitrogen fixation should be paired with efforts at obtaining strains into culture.

Diazotrophs can be selectively isolated from soil samples in N-free media because of their ability to use N_2 from air (Badalamenti et al. 2016; Ding et al. 2005; RoÂzÇycki et al. 1999). Since nitrogenase is sensitive to O_2, the culture conditions are usually anaerobic or microaerobic, and media formulations that include copiotrophic and oligotrophic conditions with or without Mo or V (Khadem et al. 2010). Usually diazotrophs are grown in liquid and semisolid media to maintain reduced oxygen condition, but Mirza and Rodrigues developed and evaluated solid media to grow diazotrophs (Mirza and Rodrigues 2012). Burk's nitrogen free medium (Park et al. 2005), nitrogen free malate medium (Dobereiner, Marriel, and Nery 1976), Norris glucose nitrogen free medium , Burke's Modified Nitrogen Free medium, and modified Rennie medium (Atlas 2010) have been used for isolation of diazotrophs.

1.1.5.2 Culture independent methods

Although isolation based studies have been extensively used for the study of diazotrophic communities it is prone to isolation of oligotrophic bacteria and fungi and a vast majority of diazotrophs has very specific conditions required for diazotrophy which is difficult to provide in the laboratory (Rosenberg et al. 2013). Various culture independent methods like polymerase chain reaction (PCR), denaturing/temperature gradient gel electrophoresis (DGGE or TGGE) (G. Muyzer, de Waal, and Uitterlinden 1993) and restriction fragment length polymorphisms (RFLPs) and terminal-RFLPs (Schütte et al. 2008) have been used extensively for culture independent study of

diazotrophs. With the advent of next generation sequencing (NGS) techniques DNA/RNA sequencing methods have been accessible for the study of diazotrophic communities (Shendure and Ji 2008).

PCR of the marker gene *nifH* has been extensively used for the study of diazotrophic communities in a wide range of environments including marine (Kapili et al. 2020), terrestrial (Tu et al. 2016), extreme (Desai, Assig, and Dattagupta 2013), anthropogenic (Héry et al. 2005), host-associated (Yamada et al. 2007) and agricultural settings (Rilling et al. 2018). Zehr and McReynolds developed the first published primers for amplification of the *nifH* gene, and these were highly degenerate because of the highly variable 5' end in the *nifH* gene. Since then, several universal and group specific *nifH* primers have been developed. Gaby and Buckley did the *in silico* and empirical analysis of the coverage of 51 universal and 35 group-specific *nifH* primers and found IGK3/DVV to perform best in both *in silico* and empirical testing. However, they also reported that these universal primers are prone to amplify homologs *nflH* and *bch/chlL* genes so, an approach to amplify total *nifH* sequences from environmental samples should include filtering-out the homologous sequences before the OTU clustering.

1.2 Prairie field and nitrogen fixation in grassland ecosystem

Grasslands are biological communities containing few trees and characterized by mixed herbaceous vegetation dominated by grasses. It is usually too dry to be a forest, and too wet to be a desert. Grasslands are found all over the globe. Dense bamboo grasslands are found in the tropics, steppes in Eastern Europe, arctic grasslands on the northern region and dry plains in central America. However, all these grasslands have

some common abiotic and biotic features (Risser 1985). Prairies are vast grassland occupying the region from central America extending North from Texas to Manitoba. (Weaver 1954). It is a part of the arid/semiarid grassland, and makes up approximately 6.9% of the earth's ice-free land, and 21% of land in the United States (Hartemink and Bockheim 2013). Most of this landscape has been appropriated by human beings as agricultural land, but the remaining portions are still protected in several conservation areas and are habitat to diverse plants, animals, and microbial lifeforms.

Different types of grasses are found in prairie. Sod-forming grasses reproduce seeds, but also produce underground rhizomes which make dense networks of shoots like in a field of slough grass or Switchgrass, or western wheatgrass. Sod forming grasses can rapidly and completely occupy the soil, creating a unique biome for soil bacteria. Also, based on the height these grasses can attain, they can be divided into three groups. Tall grasses like switch grass, slough grass, and big bluestem can attain a height of 1.5 to 2.5 meters. Prairie with limited supply of water has grasses of medium height ranging from 0.6 to 1.2 meters, and in the drier regions short grasses like blue grama, hairy grama with average height of 0.15 to 0.5 meters are dominant (Risser 1985).

In addition to the grasses, the surface of grasslands is covered by dead and decaying organic matter, and microbial populations, which comprise the biological soil (BSC). BSCs are the major source of nitrogen in arid and semiarid lands (Belnap 2002) . In the areas with minimum human activities, microorganisms present in BSC play a vital role in fixation of carbon and nitrogen (Steven et al. 2015). Free-living diazotrophic bacteria contribute to substantial amounts of nitrogen in grassland ecosystem where symbiotic

nitrogen fixation is rare (Taylor, Chazdon, and Menge 2019). A study of asymbiotic nitrogen fixing microorganism in natural, sown, and partially degraded alpine grassland on the Tibetan Plateau identified several diazotrophic bacteria from the phylum Proteobacteria, and observed significant differences in diversity of asymbiotic nitrogen fixing bacteria in these grasslands (Li et al. 2021). Similarly, a study of nitrogen fixing bacteria associated with Switchgrass in native tallgrass prairie of north Oklahoma reported several species of Alphaproteobacteria and Firmicutes having *nifH* genes. Although other studies (Davis et al. 2010; Patra et al. 2006) have also shown that nitrogen fixation in soil is significantly affected by the grass species abundant in the given grassland, the physiology and limitations of free-living diazotrophs in natural grassland ecosystems are virtually unexplored and this signifies the need for further study (Vitousek et al. 2013).

1.3 Research questions and specific objectives

After the thorough examination of literature on diversity, mechanism, and physiology of biological nitrogen fixation, I identified that the phylogeny of diazotrophs has not been on par with the growing whole genome database. Literature on diazotrophs is replete with reports of diazotrophy in multiple novel taxa but often lacks concrete biochemical and genomic evidence. Therefore, there is impending demand of a holistic approach to link these reports to available genomic data. Similarly, natural grassland is one of the least explored niches regarding the diversity and capability of biological nitrogen fixation, especially free-living nitrogen fixing bacteria. Therefore, this study addressed the above issues with the following specific objectives:

i. To perform an updated phylogenetic study of nitrogenase enzyme from all the diazotrophs.

ii. To study the diversity of free-living nitrogen fixing bacteria in a natural grassland ecosystem by culture-based approach.

iii. To characterize the nitrogen fixation by the isolates obtained above by qualitative and quantitative methods.

iv. To study the nitrogen cycling in the natural grassland ecosystem by meta-t-ranscriptomic approach.

Chapter 2

Phylogeny of Nitrogenase Structural and Assembly Components Reveals New Insights into the Origin and Distribution of Nitrogen Fixation across Bacteria and Archaea

Original Document:

Koirala, Amrit, and Volker S. Brözel 2021. "Phylogeny of Nitrogenase Structural and Assembly Components Reveals New Insights into the Origin and Distribution of Nitrogen Fixation across Bacteria and Archaea" *Microorganisms* 9, no. 8: 1662. https://doi.org/10.3390/microorganisms9081662

2.1 Introduction

The nitrogen biogeochemical cycle requires reduction of atmospheric nitrogen to ammonia. About half of the 413 Tg reactive nitrogen introduced annually to the biosphere is derived through biological nitrogen fixation by the prokaryotic nitrogenase enzyme complex, with 140 $Tg.yr^{-1}$ fixation in marine environments and 58 $Tg.yr^{-1}$ through non-agricultural terrestrial fixation, and only 2% (5 $Tg.yr^{-1}$) contributed by lightning (Fowler et al. 2013). The other half is contributed through a combination of synthetically fixed nitrogen and agricultural promotion of bacterial fixation in legumes. Biological nitrogen fixation or diazotrophy probably evolved 3.6–3.2 Ga ago to support expansion of biota in the nitrogen poor environment of that era (Stüeken et al. 2015; Weiss et al. 2016). Today it is the second most vital process for life on earth after photosynthesis (Navarro-González, McKay, and Mvondo 2001). Diazotrophs are divided into two major types: (1) symbiotic nitrogen-fixing bacteria which form symbiotic relation with legumes like *Rhizobium*, with actinorhizal plants such as *Frankia*, and

Cyanobacteria associated with cycads, and (2) free-living nitrogen fixers belonging to genera such as *Azotobacter* and *Clostridium* (Postgate et al. 1982).

Irrespective of their lifestyle all diazotrophs have a nitrogenase multi-subunit enzyme complex which is the only known natural system that catalyzes the breakdown of the triple bond between two nitrogen atoms in N_2 (Kim and Rees 1994). This oxygen-sensitive nitrogenase complex has evolved into three variants; molybdenum nitrogenase (Nif), the rarer vanadium nitrogenase (Vnf), and iron-only nitrogenase (Anf) (Hu and Ribbe 2015). Vnf and Anf nitrogenase are also known as alternative nitrogenase. An oxygen insensitive nitrogenase was reported in *Streptomyces thermoautotrophicus* (Gadkari et al. 1992), but this has recently been refuted (MacKellar et al. 2016). Molybdenum nitrogenase encoded by the *nifH*, *nifD*, and *nifK* genes are the most prevalent nitrogenase. Nitrogenase and its homologs have been classified into five phylogroups based on NifH and NifD proteins and their homologs by Raymond et al. (Raymond et al. 2004). Group I contain Mo-dependent nitrogenase from aerobic and facultative bacteria and II contains the Mo-dependent nitrogenase from anaerobic bacteria. Group III contains both the alternative forms of nitrogenase. Groups IV and V contain uncharacterized nitrogenase homologs and chlorophyll/bacterio-chlorophyll biosynthesis genes, respectively. Even though Nif homologs in Groups IV and V were thought to be unable to reduce nitrogen, one report suggested that *Endomicrobium proavitum*, which encodes a type IV nitrogenase, can fix nitrogen (Zheng et al. 2016).

NifH is the most sequenced and studied of the three core nitrogenase components, NifH, NifD, and NifK (Gaby and Buckley 2012b). Since development of the first *nifH*

PCR primers by Zehr and McReynolds in 1989, the number of complete *nifH* sequences has skyrocketed from 1500 (Zehr et al. 2003) to more than 8000 sequences at present, available in a curated NifH database (https://wwwzehr.pmc.ucsc.edu/nifH_Database_Public/). NifH phylogeny conducted by Zehr et al. in 2003 is one of the most comprehensive and is in close agreement with other NifH phylogenies (Chien and Zinder 1996; Young 2005). They classified NifH into four clusters. Cluster I consist of Mo-containing *nifH* and some *vnfH*, Cluster II consists of *anfH* and some *nifH* from Archaea, and Cluster III consists of *nifH* sequences from a diverse group of anaerobic bacteria. Cluster IV consists of *nifH* homologs, some of which are uncharacterized *nif*-like sequences and others are chlorophyllide reductase genes. According to a cross-system comparison of *nifH* diversity using 16,989 publicly available *nifH* sequences, Cyanobacteria and the Proteobacteria are the most common taxa containing a *nifH* gene (Gaby and Buckley 2011). These authors also reported that soil has the most diverse *nifH* sequences compared to marine or mat habitats.

Molybdenum nitrogenase is found in all diazotrophs and is a complex enzyme with two components. Component I, or Dinitrogenase, is a heterotetramer of NifD and NifK proteins which contains an iron-molybdenum cofactor (FeMo-co) in the active site of NifD. The second component, or Dinitrogenase reductase, is a NifH homodimer (Andrew Jasniewski et al. 2018) (Figure 2.1). In addition to these structural proteins, several ancillary proteins are required for diazotrophy, including but not limited to proteins for FeMo-co biosynthesis (NifENBXUSVYQ), nitrogenase maturation (NifZM), and *nif* gene expression activation (NifA) (Rubio and Ludden 2008). In some diazotrophs, NifA activity is controlled by the anti-activator NifL (Poza-Carrión,

Echavarri-Erasun, and Rubio 2015b), so the list of accessory proteins varies according to the species and physiological condition under which it fixes nitrogen. A study of 2000 complete genomes available in 2012 led to proposing NifHDKENB (Figure 2.1) as the minimum criteria for computational prediction of diazotrophy (Dos Santos et al. 2012). This six gene criterion has been widely used as diagnostic for diazotrophy in culture-independent studies (Delmont et al. 2018; Higdon, Pozzo, et al. 2020; Kruse et al. 2019; Meher et al. 2018).

The most updated phylogenies of nitrogenase are from 2011 based on NifH only (Gaby and Buckley 2011), and 2013 using NifHDK (Boyd and Peters 2013), and introduction of multiple *nifH* primers has caused a surge of *nifH* sequences which in turn has increased the tendency to assign diazotropy to new species without strong biochemical evidence. This has led to a significant deviation from the gold standard of biochemical evidence and caused potential incorrect assignment of phylogeny to new sequences. Hence there is an acute need for a rigorously curated *nif* database based on high quality sequences obtained from the complete genomes of cultured isolates. We exploited the recent surge in the number of species with fully sequenced genomes (28,483) to conduct a phylogenetic study of diazotrophs using the six core Nif proteins H, D, K, E, N, and B from publicly available complete genomes. Nif genes were assigned to taxa using a reference phylogeny of bacterial and archaeal genomes, based on a comprehensive set of 381 marker genes (Q. Zhu et al. 2019) to present a clear view of distribution of diazotrophic species among different prokaryotic phyla. This holistic approach also provides a curated database of six Nif proteins and identifies several genera hitherto not known to fix nitrogen but with potential to do so. The higher

resolution of this phylogenetic analysis enabled us to track the evolution and horizontal gene transfer of nitrogenase in individual phyla.

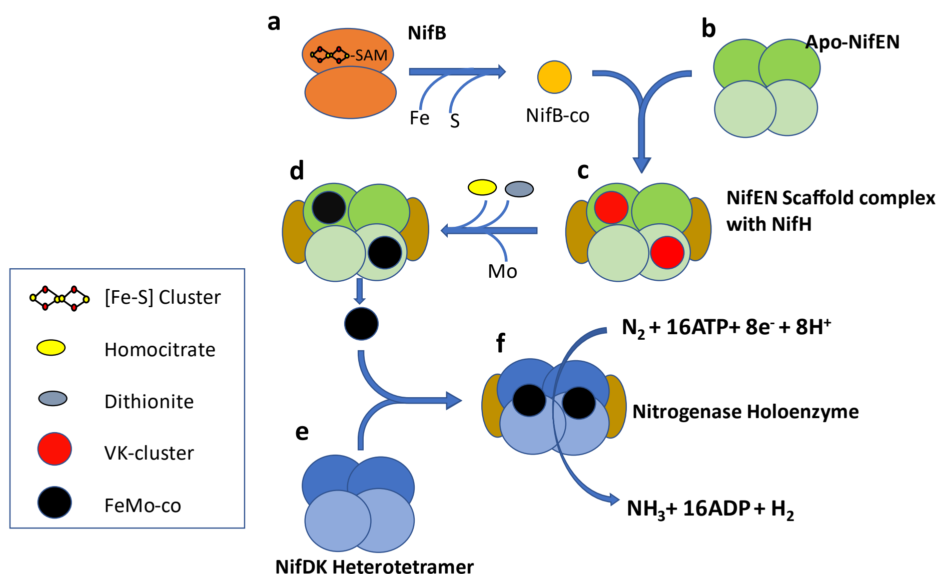


*Figure 2. 1 Schematic representation of FeMo-co biosynthesis. NifB (**a**) is a homodimer and belongs to the SAM radical protein family. NifB supplies homocitrate and Mo free, Fe-S cluster (precursor of Fe-Mo cofactor) to the Apo-NifEN heterotetramer (**b**). NifEN is homologous to NifDK and acts as molecular scaffold for assembly of Fe-Mo cofactor along with NifH (**c**,**d**). The assembled Fe-Mo cofactor is transferred to the Nitrogenase apoenzyme (**e**), which forms the nitrogenase holoenzyme (**f**) along with NifH. NifH has a role in both assembly of the Fe-Mo cofactor as well as reduction of dinitrogen (Burén et al. 2020).*

2.2 Materials and Methods

Annotated NifH, NifD, NifK, NifE, NifN, NifB, VnfH, VnfD, VnfK, AnfH, AnfD, and AnfK protein sequences were downloaded from the InterPro database (www.ebi.ac.uk/interpro/) (April 2019). Since InterPro is an integrated database built upon the signatures from several member databases like Pfam, PANTHER, and TIGRFAMs, annotation in InterPro is more reliable than individual databases. Sequences were grouped together by taxon name and only genome sequences having all the six genes (*nif*HDKENB) were selected for further study. In cases where there is more than

one copy of *nif* gene per organism, only that *nif* gene occurring in the same genomic neighborhood of the remaining minimal *nif* gene complement was manually selected to yield a single set of *nif*HDKENB sequences for each organism. Organisms with fused *nif*EN and *nif*NB were also identified based on their sequence length and included. The genomes of organisms containing the defined *nif* genes were obtained from NCBI (www.ncbi.nlm.nih.gov) using CDS IDs obtained from mapping function in uniport (www.uniprot.org). Protein sequences are included as Supplemental Data S3 (https://www.mdpi.com/2076-2607/9/8/1662/s1).

Biosample IDs associated with the genomes in NCBI were used to obtain the attributes describing sample type (cultured or metagenome), source of isolation, assembly status, biotic relationship (free-living or host associated), host (if applicable), env_material, and env_feature. These attributes were used to get information on the habitat and lifestyle adopted by diazotrophs. All pertinent information is included as Supplemental Data S1 and S2 (https://www.mdpi.com/2076-2607/9/8/1662/s1).

Protein sequences were aligned individually using ClustalW multiple sequence alignment (version 2.1) in Galaxy (https://usegalaxy.org/) with default parameters and concatenated in R using the Phylotools package (version 0.2.2). Phylogenetic trees from the concatenated sequences were constructed using PhyML 3 (Guindon et al. 2010), FastTree V2.1.10 (Price, Dehal, and Arkin 2010) using the JTT+CAT evolution model, and RapidNJ (Simonsen, Mailund, and Pedersen 2008) using the Kimura model. For both FastTree and RapidNJ default amino acid substitution model was used in the final analysis as other models also produced similar tree topologies. Branch support for

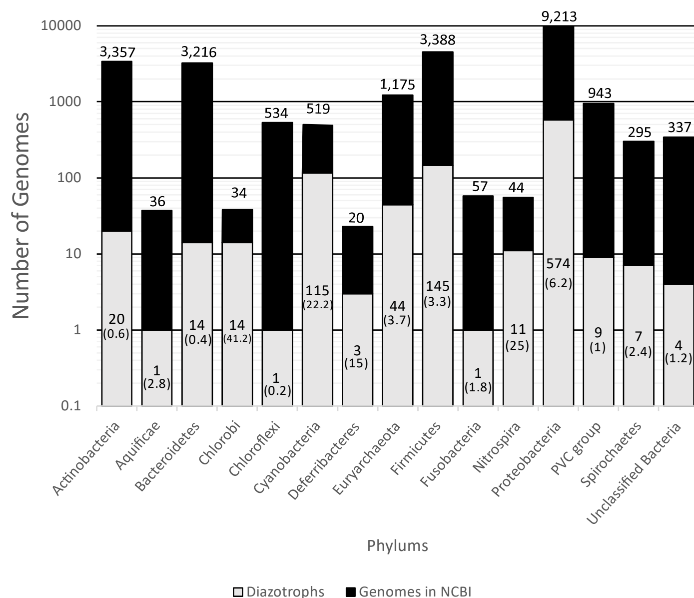
PhyML 3 was calculated using a Bayesian-like transformation of aLRT (aBayes) (Anisimova et al. 2011) as it was the only method computationally feasible to run at HPC facility. Phylogeny of NifHDKENB was also compared by combining individual Nif trees as well. ASTRAL-III (Zhang et al. 2018) was used to obtain a combined tree from six individual Nif trees obtained by using FastTree using the JTT+CAT evolutionary model. The ASTRAL tree was compared with concatenated trees by forming a consensus tree of three concatenated trees by using consensus clustering of phylogenetic trees obtained using Rphylip (Revell and Chamberlain 2014) which is an R implementation of Phylip (Felsenstein 1993). Trees were annotated using iTOL (Letunic and Bork 2019) and Fig Tree v1.4 (Rambaut 2012).

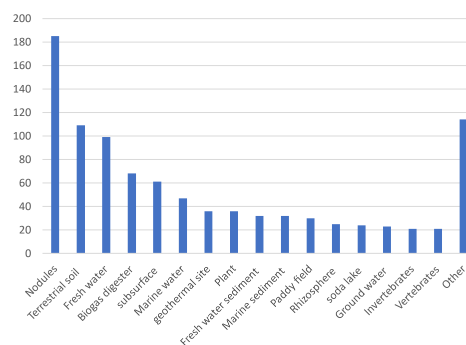
The 16S rRNA sequences were obtained from genomic assemblies in NCBI using the annotation key words 16S rRNA or 16S ribosomal RNA. The tree of life proposed by Zhu et al. (Q. Zhu et al. 2019) was used to delineate distribution of diazotrophs across taxa and to estimate geological time for the evolution of diazotrophy in different phyla.

2.3 Results

In total, 963 genomes or genomes from metagenomes containing all six *nif* genes were obtained, falling into 325 genera from Actinobacteria, Aquificae, Bacteroidetes, Chlorobi, Chloroflexi, Cyanobacteria, Deferribacteres, Euryarchaeota, Firmicutes, Fusobacteria, Nitrospira, Proteobacteria, PVC group, and Spirochaetes (Figure 2.2). Of the 24,168 genomes available in NCBI (April, 2020), *nif* genes were found in limited proportion, 0.2 to 41.1% per phylum, showing the wide but patchy distribution of diazotropy in both bacterial and archaeal phyla (Figures 2 and 3). Most organisms were

from Proteobacteria, followed by Firmicutes and Cyanobacteria. Some smaller phyla like Aquificae, Chloroflexi, and Fusobacteria had only one representative each. *Mesorhizobium* with 66 strains was the most sequenced diazotroph followed by *Rhizobium* (44), and *Bradyrhizobium* (41) (Figure 2.3).





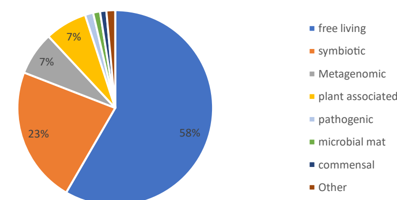


*Figure 2. 2 Distribution, habitat, and lifestyle of diazotrophs in various bacterial and Archaeal phyla. (**a**) The black bar represents the total number of genomes sequenced, and the gray bar the total number of diazotrophs in that phylum. Labels on top of black bar indicate the total number of sequenced genomes. Labels in the gray bar represent the number of genomes containing nifHDKENB in that phylum (numbers in parentheses are the percentage of sequenced genomes that have nifHDKENB). (**b**) Distribution of organisms having all six nif genes by habitat and (**c**) pie diagram showing the lifestyle adopted by various diazotrophs.*

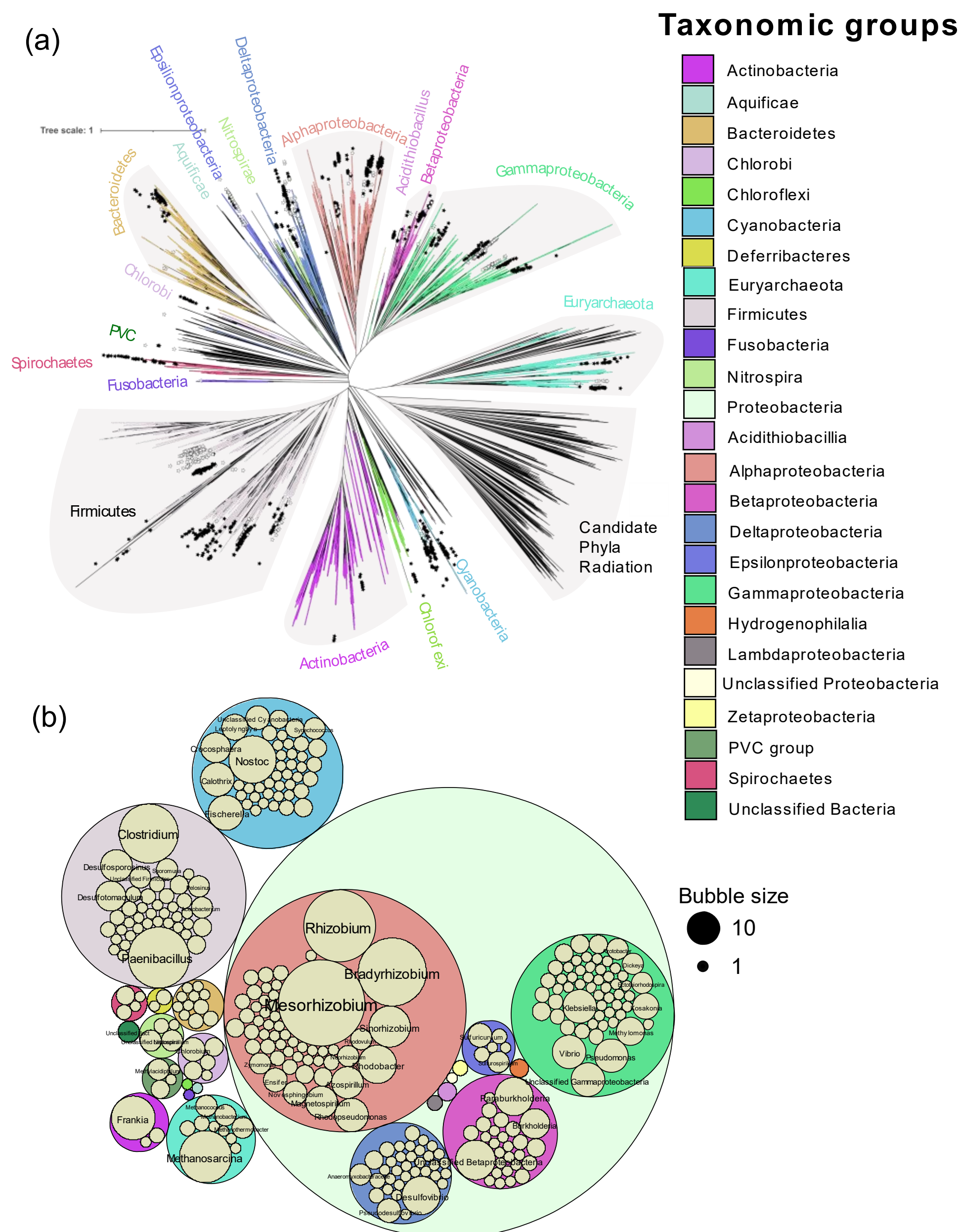


*Figure 2. 3 Distribution of diazotrophs by genus. (**a**) Distribution of diazotrophs in the tree of life proposed by Zhu et al. (Q. Zhu et al. 2019). Nexus tree file for all prokaryotes was visualized in iTOL and leaves were annotated by genus according to the evidence of diazotrophy observed in this study. Solid and open symbols represent genera with and without biochemical evidence respectively. Genera without a star had neither genomic nor biochemical evidence of diazotrophy. (**b**) Distribution of the diazotrophic genera by the number of species observed in each genus. Bubble size corresponds to the number of assemblies of that genus and phylum.*

Most of the genomes (867) were from cultured isolates, followed by metagenomic assembly (93), and single cell genomes (3). Diazotrophic organisms were from a wide range of habitats (Figure 2.2). The largest proportion of diazotrophs were isolated from root nodules (180), followed by terrestrial soil (109), fresh water (99), and biogas digesters (68). More than half (562) of the organisms were free living followed by a symbiotic lifestyle. Most of the symbiotic bacteria were rhizobia (165), fixing nitrogen in root nodules of leguminous plants. Other modes of symbiosis observed were root nodules in actinorhizal plants by *Frankia* sp. (15), cyanobacterial endosymbionts of diatoms (3), syntrophic coculture (*Syntrophobotulus glycolicus*, *Chlorobium ferrooxidans*), phototrophic consortium (*Chlorobium chlorochromatii*), Proteobacteria from gill tissue of bivalves (5), Spirochaetes from termite gut (3), *Nostoc* in coralloid roots (3), moss carpet (5), and Azolla (1). Thirty-six strains were isolated from thermophilic sites and eight strains isolated from psychrophilic sites (Supplementary Data S2) (https://www.mdpi.com/2076-2607/9/8/1662/s1).

2.3.1. Biochemical Evidence of Diazotrophy

Incorporation of $^{15}N_2$ and the acetylene reduction assay have been used to biochemically confirm diazotrophy. With the growing ease of genome sequencing, isolates can be tentatively reported to fix nitrogen based on occurrence of a *nif* operon in their genome. Of the 325 genera containing *nif*HDKENB genes, 156 were without biochemical evidence of diazotrophy and most of them were represented by single strains (Supplementary Data S3) (https://www.mdpi.com/2076-2607/9/8/1662/s1). Figure 2.3 shows the distribution of prokaryotic genera with and without biochemical evidence in a

tree of life proposed by Zhu et al. (Q. Zhu et al. 2019). No biochemical evidence was found in the phyla Aquificae and Fusobacteria, both of which are represented here by a single strain. Most of the isolates in the delta and epsilon subdivisions of the Proteobacteria do not have biochemical evidence of N_2 fixation to date.

2.3.2. Phylogeny of Nitrogenase

A total of 6 genes were obtained from all 963 organisms, but 29 assemblies from Cyanobacteria and Nitrospira had fused *nifE* and *nifN*, and 44 assemblies from Clostridia which had fused *nifN* and *nifB*. In addition to the Mo-Fe nitrogenase, 31 assemblies had vanadium containing V-Fe nitrogenase and 44 assemblies had iron only containing Fe-Fe nitrogenase. Species of *Azotobacter*, *Clostridium*, *Methanosarcina*, *Paenibacillus*, and *Rhodopseudomonas* were found to harbor all three forms of nitrogenase. Not all alternative nitrogenases were found to carry their own cofactor synthesis genes (*nif*ENB), hence phylogenetic analysis with six genes was limited to the Mo-Fe nitrogenase only. Phylogenetic analysis of Mo-Fe nitrogenase was done using multiple approaches. Reconstruction of species trees from six individual Nif protein trees (Supplementary Figures S1–S6 https://www.mdpi.com/2076-2607/9/8/1662/s1) using ASTRAL-III gave similar patterns of clustering to the tree constructed using concatenated sequences. Trees constructed with multiple algorithms (PhyML, FASTTREE, and Rapid NJ) from concatenated sequences also agree on the major clusters (Supplementary Figures S7–S9 https://www.mdpi.com/2076-2607/9/8/1662/s1).

Phylogeny of the Mo-Fe nitrogenase using concatenated NifHDKENB obtained by FASTTREE is given in Figure 2.4 and the clusters are labeled according to Raymond

et al. (Raymond et al. 2004). Cluster I contains aerobic and facultative bacteria from Proteobacteria, Cyanobacteria, Firmicutes, Actinobacteria, and other smaller phyla Aquificae, Deferribacteres, and Nitrospira. Cluster II consists of anaerobic bacteria (Firmicutes, Bacteroidetes, Chlorobi, Chloroflexi, Fusobacteria) and Archaea (Euryarchaeota). Cluster III consists of Mo-Fe nitrogenase from the Methanomada clade of Euryarchaeota and alternative nitrogenase as evident from the NifHDK tree (inset in Figure 2.4). Cluster III in the *nif*HDKENB tree appears small compared to *nif*HDK because alternative nitrogenases were not included as cofactor biosynthesis proteins were not universally present in alternative nitrogenase operons.

2.3.3. Distribution and Phylogeny of Nitrogenase in Various Phyla

2.3.3.1. Proteobacteria

Proteobacteria is the largest bacterial phylum and consists of bacteria with diverse morphology and physiology yet are united by 16S rRNA phylogeny. It had the largest number of representatives in this study as well (574). The nitrogenases of Proteobacteria occur entirely in Cluster I, except for the Deltaproteobacteria whose nitrogenases cluster together with other anaerobic bacteria in Cluster II.

Alphaproteobacteria are distributed in Clusters IA and IB (Figure 2.5). The orders Magnetococcales, Rhodospirillales, Rhodobacterales, Sphingomonadales, and Rhizobiales were found to contain nif genes. Cluster IA is coherent with 16S rRNA phylogeny of Alphaproteobacteria where primitive Rhodospirillales, Rhodobacterales, and Sphingomonadales branch out earlier than Rhizobiales. Rhizobiales are distributed between Cluster IA and IB. Phyllobacteriaceae, Rhizobiaceae, and Rhodobiaceae form a

monophyletic cluster in Cluster IA whereas Bradyrhizobiaceae, Xanthobacteraceae, and Beijerinckiaceae form a monophyletic cluster in IB along with Betaproteobacteria and other acidophilic methanotrophic bacteria. Some Alphaproteobacteria (*Magnetococcus*, *Magnetovibrio*, *Martelella*, and *Cohaesibacter*) occur in Cluster ID along with Gammaproteobacteria.

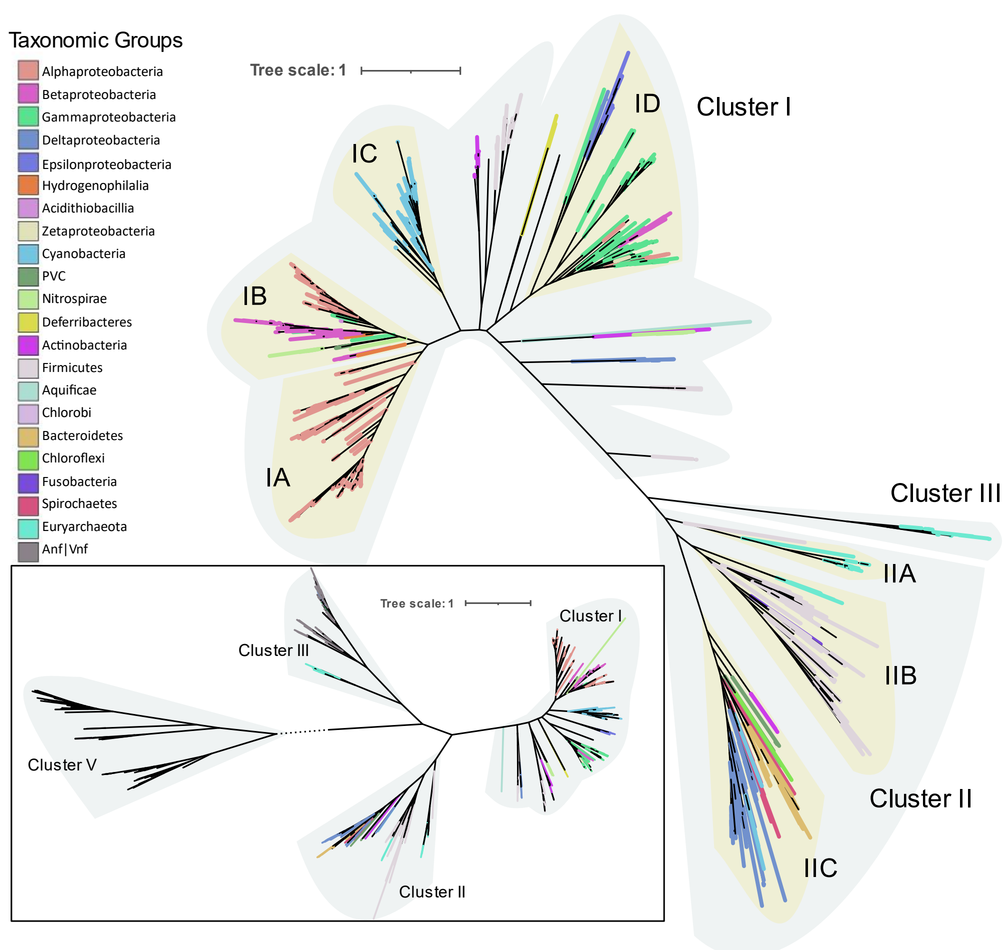


Figure 2. 4 Molecular phylogenetic analysis of concatenated NifHDKENB proteins by FastTree using the JTT+CAT evolution model. Each clade is highlighted by the bacterial or archaeal phylum and Proteobacteria are further divided into classes. Inset: Phylogenetic analysis of the NifHDK protein sequences with concatenated bclLNB as outgroup. All the clusters are labeled as suggested by Raymond (Raymond et al. 2004).

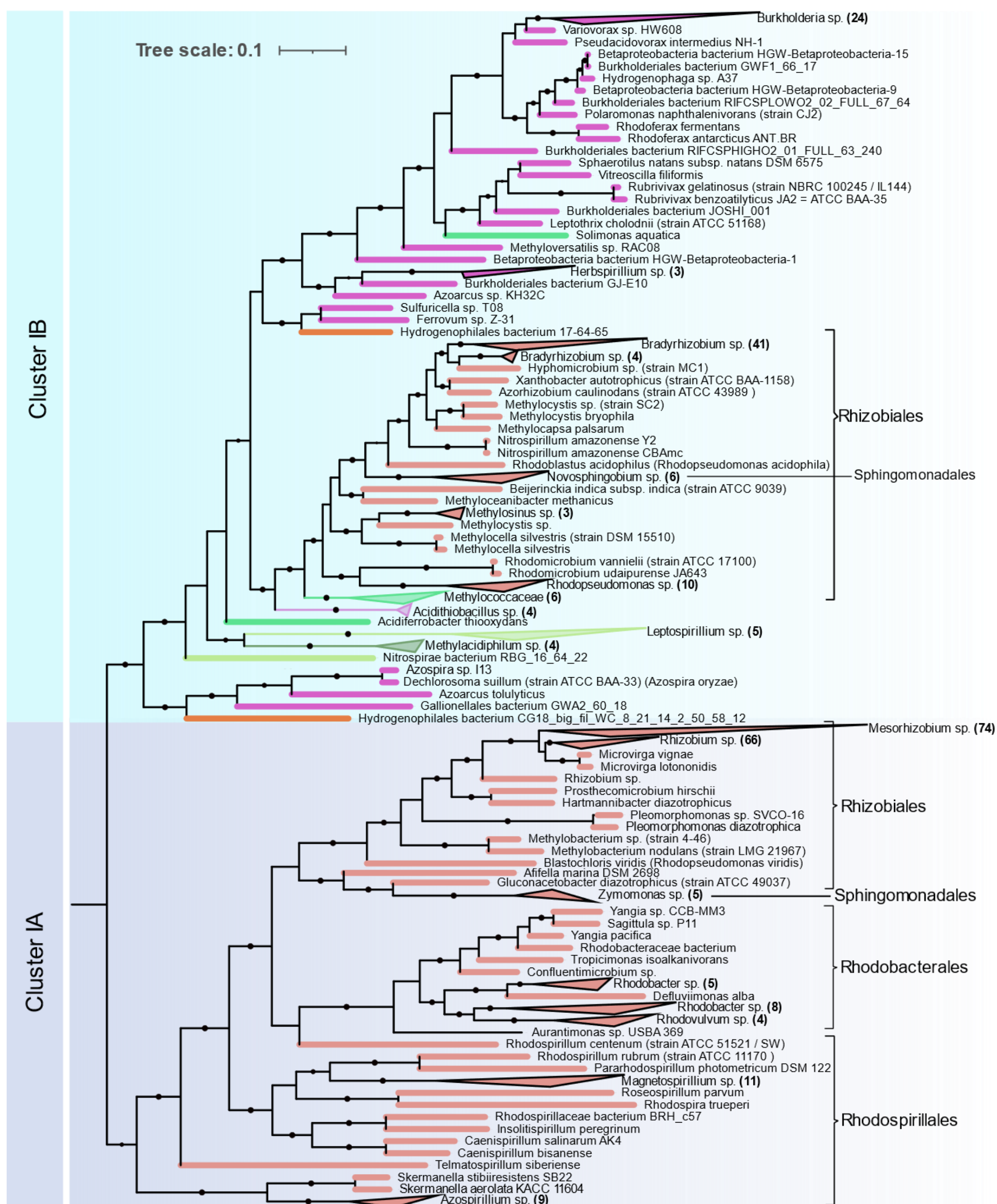


Figure 2. 5 NifHDKENB Clusters IA and IB showing the phylogenetic relationship among Proteobacteria obtained by FastTree. Branches are colored according to the Figure 2.4.

Nitrogenases from **<u>Betaproteobacteria</u>** are present in Cluster IB and are represented by the orders Ferrovales, Neisseriales, Nitrosomonadales, and Burkholderiales. The Rhodocyc

lales are present in Cluster ID. The Burkholderiales were the most abundant diazotrophic Betaproteobacteria in this study.

<u>Gammaproteobacterial</u> nitrogenases form a phylogenetically coherent cluster in ID. Enterobacterales and Pseudomonadales were the most common diazotrophic Gammaproteobacterial. Methanotrophic Methylococcales form a deep branch in the Rhizobiales cluster in IB.

Nitrogenases from **<u>Deltaproteobacteria</u>** occur in a monophyletic group in Cluster IIC except for the orders *Desulfuromonadales* and *Myxococcales* which occur at the base of cluster I. Most of the *Deltaproteobacteria* genomes in Cluster IIC are from *Desulfovibrionales* and *Desulfobacterales* and for several of these isolates, no biochemical evidence could be found (Supplement Table S3 https://www.mdpi.com/2076-2607/9/8/1662/s1). Distribution of nitrogenase in Deltaproteobacteria in Cluster IIC is also coherent with 16S rRNA phylogeny. Similarly, *Desulfuromonadales* (strictly anaerobic) and *Myxococcales* (aerobic/facultative anaerobic) that occur in Cluster I were found to be phylogenetically distinct from the rest of the Deltaproteobacteria in a recent phylogenetic study (Q. Zhu et al. 2019).

<u>Epsilonproteobacteria</u> are the next most abundant Proteobacteria represented by 15 isolates, all of which occur in a monophyletic cluster in ID along with Gammaproteobacteria.

2.3.3.2. Archaea

Diazotrophic Archaea were only found in methanogenic Euryarchaeota, but they are distributed in both Clusters II and III (Figure 2.6a). *Methanosarcina* is the most sequenced diazotrophic archaea (24 out of 44 genomes). Its nitrogenase forms a single monophyletic cluster in IIA along with other *Methanosarcinales* (*Methanothrix* and *Methanolobus*). *Methanomicrobiales* (*Methanosphaerula*, *Methanolacina*, and *Methanoregula*), and *Methanocellales* (*Methanocella*) form a monophyletic group in ClusterIIB together with *Firmicutes.* Nitrogenases from the group "Methanomada" (*Methanococcales* (*Methanococcus*), and *Methanobacteriales* (*Methanobacterium*, *Methanothermobacter*)) are different from any archaeal nitrogenase and form a deep branch in Cluster III along with the alternative nitrogenases. Nif phylogeny strictly follows the 16S rRNA phylogeny, indicating the vertical transfer of *nif* genes within Archaea.

2.3.3.3. Firmicutes

Genomes of Firmicutes (145) from Bacilli (41), Clostridia (90), Negativicutes (13), and Tissierellia (1) were found to contain all six *nif* genes. *Clostridium* and *Paenibacillus* are the most prevalent diazotrophs in Firmicutes. *Nif* phylogeny shows three distinct evolutionary patterns of nitrogenase in Firmicutes (Figure 2.6b). Although distribution of Firmicute *nif* appears to be patchy and interspersed by other taxa, within each patch 16S rRNA phylogeny is highly conserved. All Bacilli (except *Desulfuribacillus*) are present as a monophyletic cluster in Cluster I together with aerobic high G+C gram positive Actinobacteria. Paenibacillaceae (*Paenibacillus* and *Fontibacillus*), and Bacillaceae (*Anaerobacillus* and *Bacillus*) form their own distinct

cluster and *Kyrpidia* branches out early, which is in close harmony with the 16S rRNA phylogeny of the Bacilli. Only two *Bacillus* (*B. nealsonii* and *B. caseinilyticus*) were found to contain complete *nif* operons.

In Cluster IIB Firmicutes occur as a monophyletic cluster which comprises Clostridia, Negativicutes, and Tissierellia. Peptococcaceae (*Desulfosporosinus*, *Dehalobacter*, *Syntrophobotulus*, and *Desulfitobacterium*) occur at the base of Cluster I. The division of Peptococcaceae into two phylogroups is also evident in 16S rRNA phylogeny of Peptococcaceae (Vos et al. 2011).

Thermoanaerobacterium which branches deep in 16S rRNA phylogeny also has very distinct *nif* genes that branch out earlier than other Firmicutes and occur in IIA.

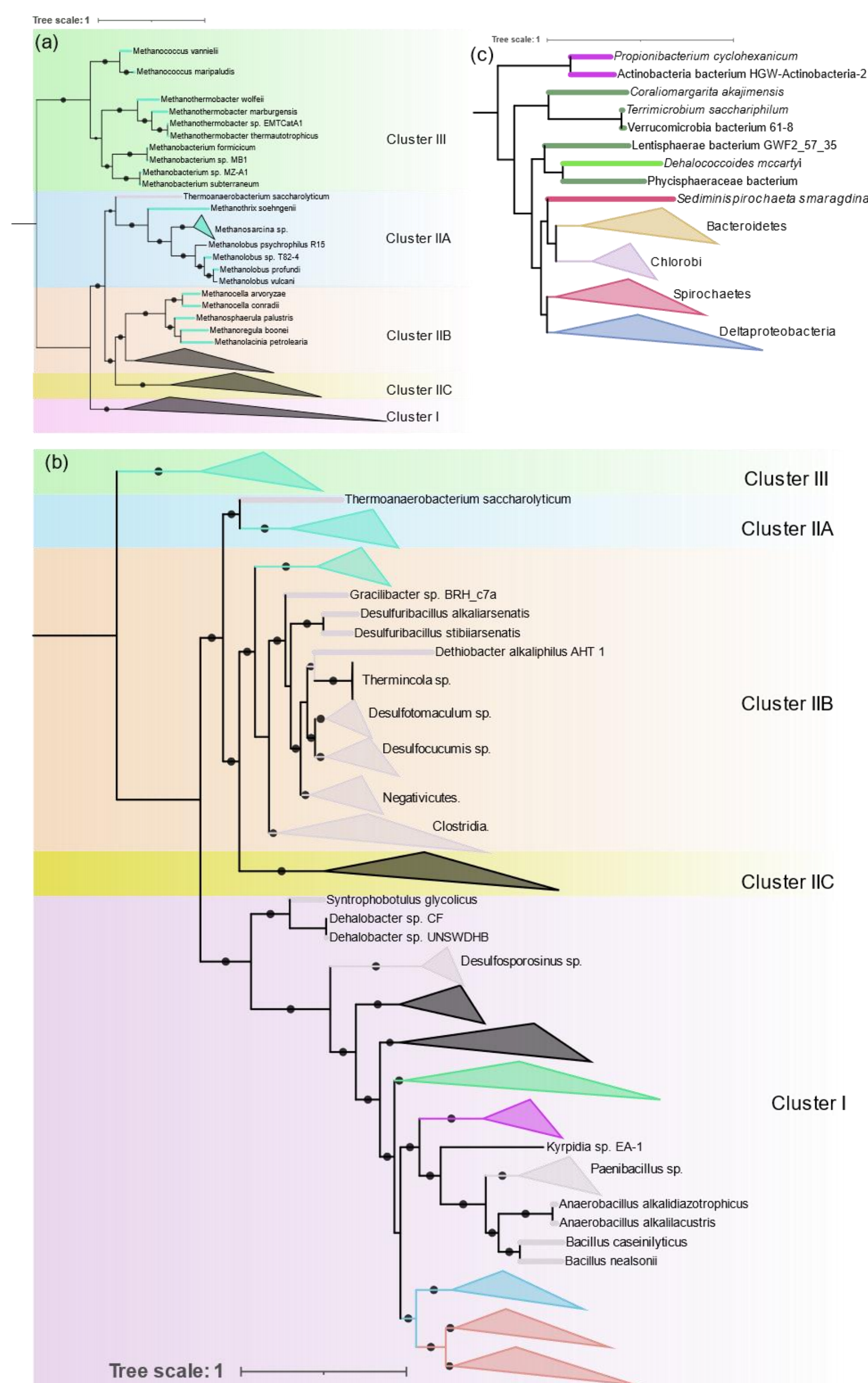


*Figure 2. 6 NifHDKENB phylogeny highlighting nitrogenase distribution in Archaea (**a**), Firmicutes (**b**), and overall distribution of Cluster IIC (**c**), obtained by FastTree, and manually rooted based on the position of the outgroup (Cluster V) in the NifHDK tree.*

2.3.3.4. Cyanobacteria

The Nif proteins of all 115 cyanobacterial genomes clustered together in a single cluster (Cluster IC) with no representatives of non-cyanobacterial taxa (Figure 2.7). This indicates a monophyletic origin of cyanobacterial *nif* genes. The *nif* phylogeny agrees with the 16S rRNA phylogeny of these cyanobacterial isolates, indicating all the cyanobacteria derived their *nif* operon from a common cyanobacterial ancestor. Cyanobacteria that possess *nif* genes were found to occur in various habitats from hot springs (*Synechococus* sp. Strain JA-2-3B'a-2-13), antartic endolith (*Chroococcidopsis* sp.), marine (*Trichodesmium erythraeum*, *Crocosphaera)* to fresh water (Oscillatoria). Within the cyanobacterial cluster these isolates form five distinct clades.

Clade I is homogenous in terms of 16S rRNA phylogeny, containing all the Nostocales. These are filamentous cyanobacteria producing specialized, terminally differentiated cells called heterocysts for nitrogen fixation. Since photosynthesis and nitrogen fixation are spatially separated from each other, these cyanobacteria are capable of fixing nitrogen in aerobic condition (Issa, Abd-Alla, and Ohyama 2014). Major representatives in this clade are *Nostoc*, *Fischerella*, *Calothrix*, *Nodularia*, and *Anabaena.*

Clade II is polyphyletic, containing Synechococcales, Oscillatoriales, Chroococcidiopsidales, and a single representative of Chroococcales, *Chroogloeocystis* where no biochemical evidence has been reported. Chroococcidiopsidales are unicellular, free-living cyanobacteria shown to fix nitrogen in anaerobic condition only (Banerjee and Verma 2009). Other cyanobacteria in this clade are all filamentous but lack heterocysts and fix nitrogen in anaerobic or microaerophilic conditions only (Stal 2015).

Clade III consists of unicellular Pleurocapsales and Chroococcales along with single filamentous genus *Lyngbya.* All Chroococcales of Cluster III form a single cluster and contain some of the most important marine diazotrophs (*Crocosphaera/*UCYN B, *Atelocyanobacterium/*UCYN A, *Gloethece*, *Rippkaea*, and *Zehria).* Chroococcales use temporal separation of photosynthesis and diazotrophy by fixing nitrogen in the dark cycle only (Muñoz-Marín et al. 2019).

Clade IV is also polyphyletic, consisting of Synechococcales, Pleurocapsales, Oscillatoriales, and Chroococcales.

Clade V contains three unique cyanobacteria characterized by deep branching from the rest of the cyanobacteria in both 16S rRNA and NifHDKENB. These are unicellular, thermophilic cyanobacteria isolated from hot springs in Yellowstone National Park (Bhaya et al. 2007). Although all three strains have all six *nif* genes, diazotrophic growth has not been reported yet.

2.3.3.5. Actinobacteria

There were 21 Actinobacterial genomes containing all 6 *nif* genes, 18 were *Frankia*, 1 *Propionibacterium*, and 2 unclassified actinobacteria. All the *Frankia* make a monophyletic cluster in Cluster I (Figure 2.6b) along with other gram-positive bacteria. Nevertheless, anaerobic *Propionibacterium* and one of the unclassified actinobacteria cluster together with *nif* from *Verrucomicrobium* in Cluster IIC (Figure 2.11). The remaining unclassified actinobacterial *nif* align *nif* from Nitrospira. All *nif* containing Actinobacteria cluster together by 16S rRNA indicating polyphyletic evolution of nitrogenase in Actinobacteria.2.3.3.6. Bacteroidetes and Chlorobi

These phylogenetically related phyla occurred as a monophyletic group in Cluster IIC (Figure 2.6c) with a separate branch for each phylum. Only 11 genera (*Bacteroides*, *Dysgonomonas*, *Paludibacter*, *Azobacteroides*, *Labilibaculum*, *Alkalitalea*, *Geofilum*, *Saccharicrinis*, *Draconibacterium*, *Lutibacter*, and *Arcticibacter*) in Bacteroidetes were found to have all *nif* genes, of which only five have been shown to fix nitrogen. Chlorobi is a small phylum of green sulfur bacteria and all the sequenced members of this phylum have *nif* genes except *Ignavibacteria.* The evolution of nitrogenase in these two phyla is in close alignment with their 16S RNA phylogeny.

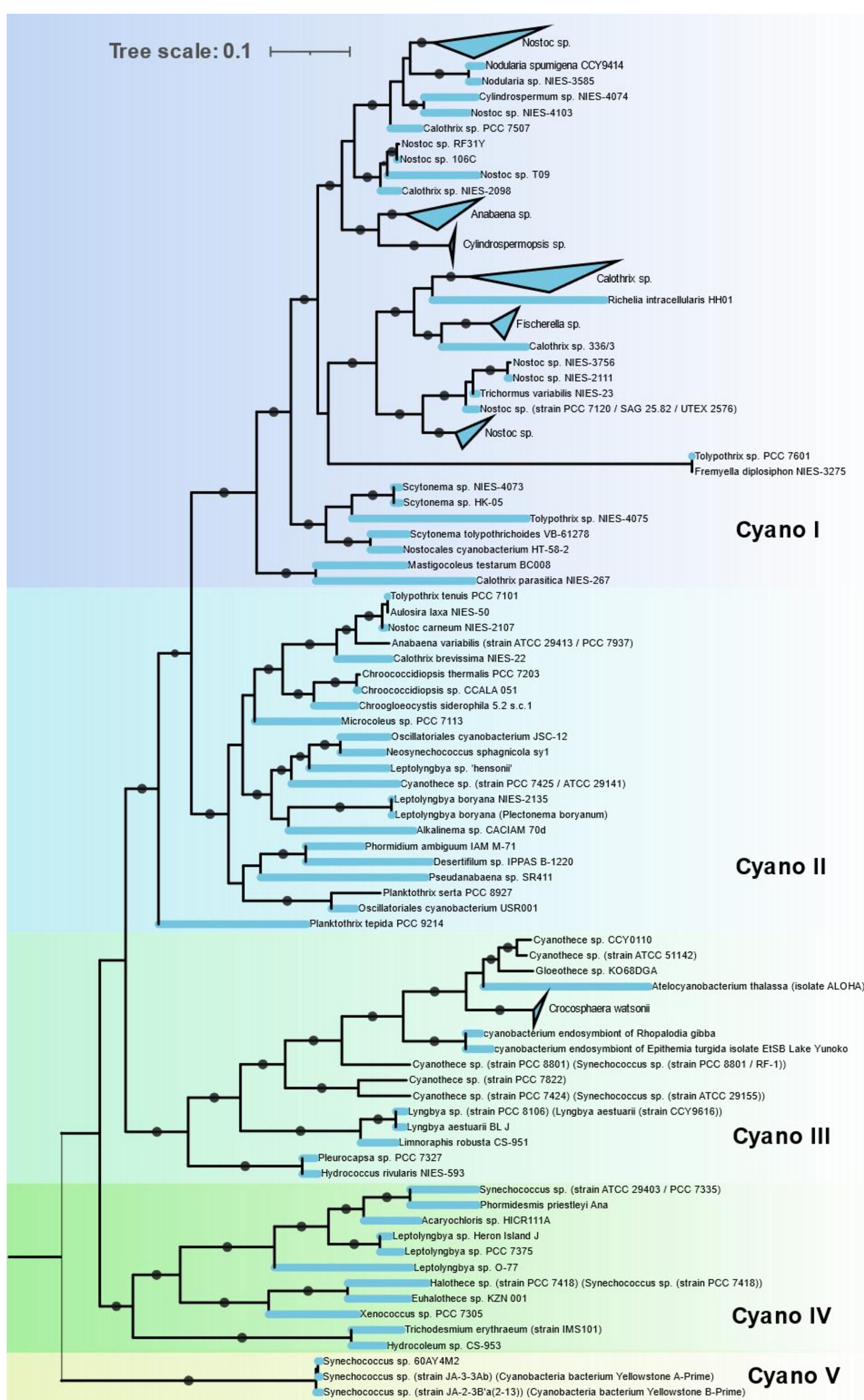


Figure 2. 7 Phylogenetic relationship among Cyanobacteria in Cluster IC of the Concatenated NifHDKENB tree obtained by FastTree.

2.3.3.7. Spirochaetes

All the Spirochaete nitrogenases form a monophyletic cluster in IIB (Figure 2.6c). All the diazotrophic Spirochaetes belonged to three genera (*Spirochaeta thermophila*, *Treponema azotonutricum*, *Treponema primitia*, and *Sediminispirochaeta smaragdinae*). All of them are strictly anaerobic.

2.3.3.8. Other Anaerobes

Verrucomicrobia, Chloroflexi, Planctomycetes, and Lentisphaerae form a single cluster in IIB (Figure 2.6c). Verrucomicrobia, Planctomycetes, and Lentisphaerae are phylogenetically related and group together as the PCV clade. One methanotrophic genus of Verrucomicrobia, Methylacidiphilum forms a separate group in Cluster IB with Nitrospira and other acidophilic Gammaproteobacteria. The Planctomycetes and Lentisphaerae representative were obtained from metagenomic sequencing. The only Chloroflexi found to harbor nitrogenase was Dehalococcoides mccartyi (D. ethenogenes) and the Fusobacteria were also represented by a single observation, Ilyobacter polytropus, both of which occur in Cluster IIB.

2.3.3.9. Nitrospira, Deferribacterales, and Aquificacea

These are small phyla represented by 10, 3, and 1 strain, respectively. Nitrospira and Aquificae form a single cluster at the base of Cluster I. Acidophilic *Leptospirillium* sp. of the Nitrospira occur at the base of Cluster IB. Deferribacterales form a single deep branch in Cluster ID.

2.3.4. Alternative Nitrogenases

Compared to the diversity of Mo-Fe nitrogenases, alternative forms of nitrogenase are limited to very few taxonomic groups as shown by phylogeny of concatenated HDK proteins. Vanadium containing nitrogenase (Vnf) is found only in some Alphaproteobacteria, Gammaproteobacteria, Firmicutes, Archaea (*Methanosarcina* sp.), and Cyanobacteria. Similarly, Fe only containing nitrogenase (Anf) is found in some Alphaproteobacteria, Gammaproteobacteria, Deltaproteobacteria, Archaea (*Methanosarcina* sp.), and Firmicutes (Figure 2.8). Phylogenetically, Vnfs and Anfs form their own distinct clades, but are very similar to Mo-Fe nitrogenases from the Methanomada clade of Euryarchaeota. Anfs and Vnfs are very similar to each other, suggesting their very recent evolution from Mo-Fe nitrogenase (inset in Figure 2.4).

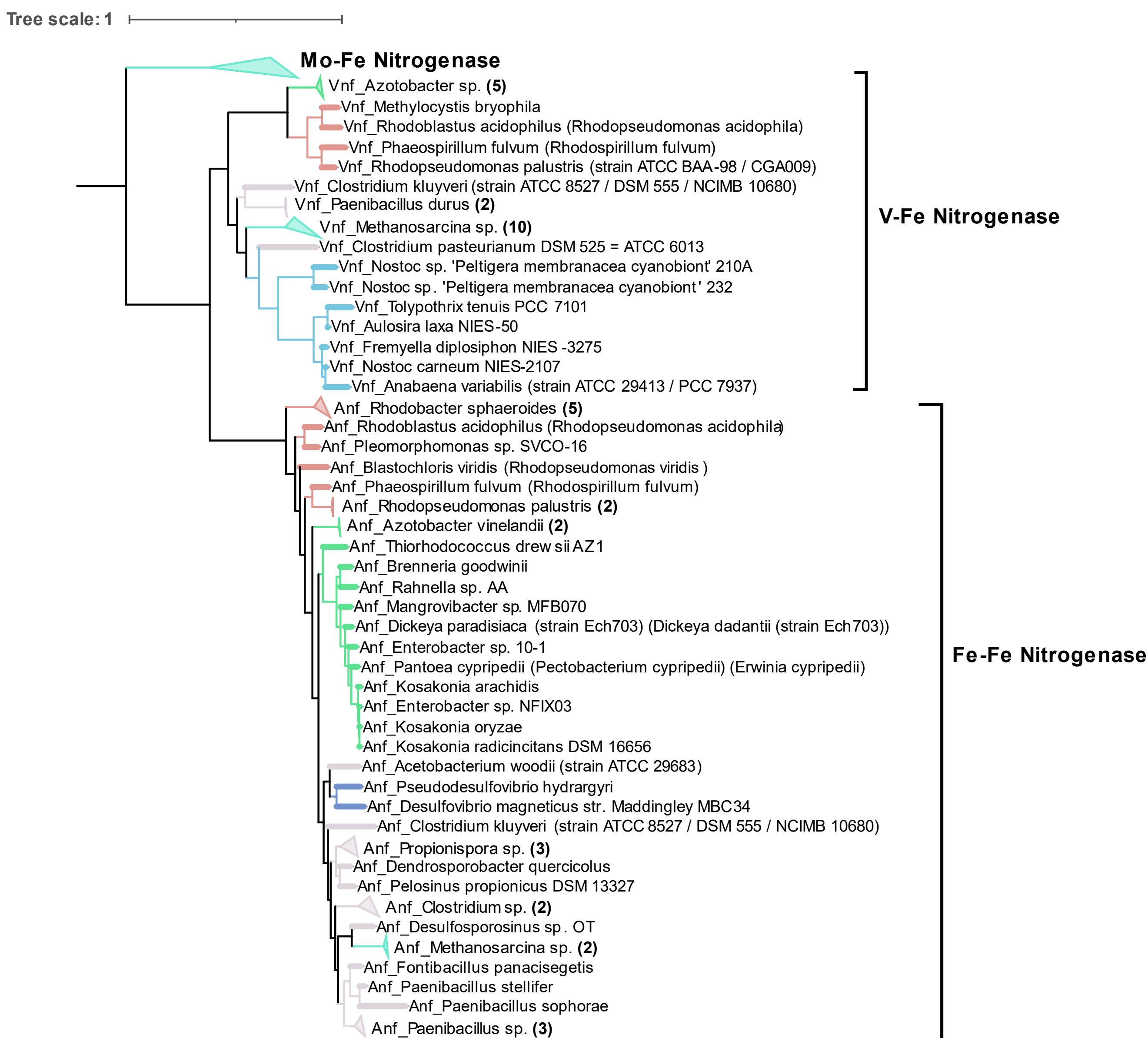


Figure 2. 8 Cluster III of the concatenated AnfHDK/VnfHDK tree obtained by FASTTREE, showing the phylogenetic relationship between isolates carrying alternative forms of nitrogenase.

2.3.5. Horizontal Gene Transfer (HGT)

Multiple instances of HGT are evident when comparing *nif* phylogeny with16S rRNA phylogeny (Figure 2.9) or the tree of life reported by Zhu et al. (Q. Zhu et al. 2019) (Supplementary Figure S10 https://www.mdpi.com/2076-2607/9/8/1662/s1). The two

most evident instances of HGT are within anaerobic niches and acidic niches where methylotrophs are concentrated. Regardless of their 16S rRNA phylogeny, the NifHDKENB sequences of strict anaerobes belonging to diverse phyla were similar, aligning together in Cluster II. This included NifHDKENB sequences from strict anaerobes like Clostridia, Bacteroidetes, Chlorobi, PVC, Chloroflexi, Fusobacteria, Deltaproteobacteria, and Spirochaetes that cluster together in Cluster II with Euryarchaea. Within Cluster II, Clostridia and other anaerobes form two clear clusters, indicating two major instances of HGT. Details are apparent when zooming in to the Supplementary Tree file (https://www.mdpi.com/2076-2607/9/8/1662/s1).

Cluster IB represents another instance of HGT where acidophilic and methanotrophic bacteria cluster together irrespective of their phylogeny. At the root of this cluster are *Methylacidiphilum* sp. from PVC, *Leptospirillium* sp. from Nitrospira, *Acidithiobacillus* from Proteobacteria, and species of Methylococcaceae from Gammaproteobacteria, all of which might have obtained their *nif* genes from ancient Alphaproteobacteria. In addition to these, there are also several instances where one or more genera appears to cluster with phylogenetically unrelated bacteria (Figure 2.9, Supplementary Tree file https://www.mdpi.com/2076-2607/9/8/1662/s1) indicating promiscuous lateral transfer during the early divergence of bacteria.

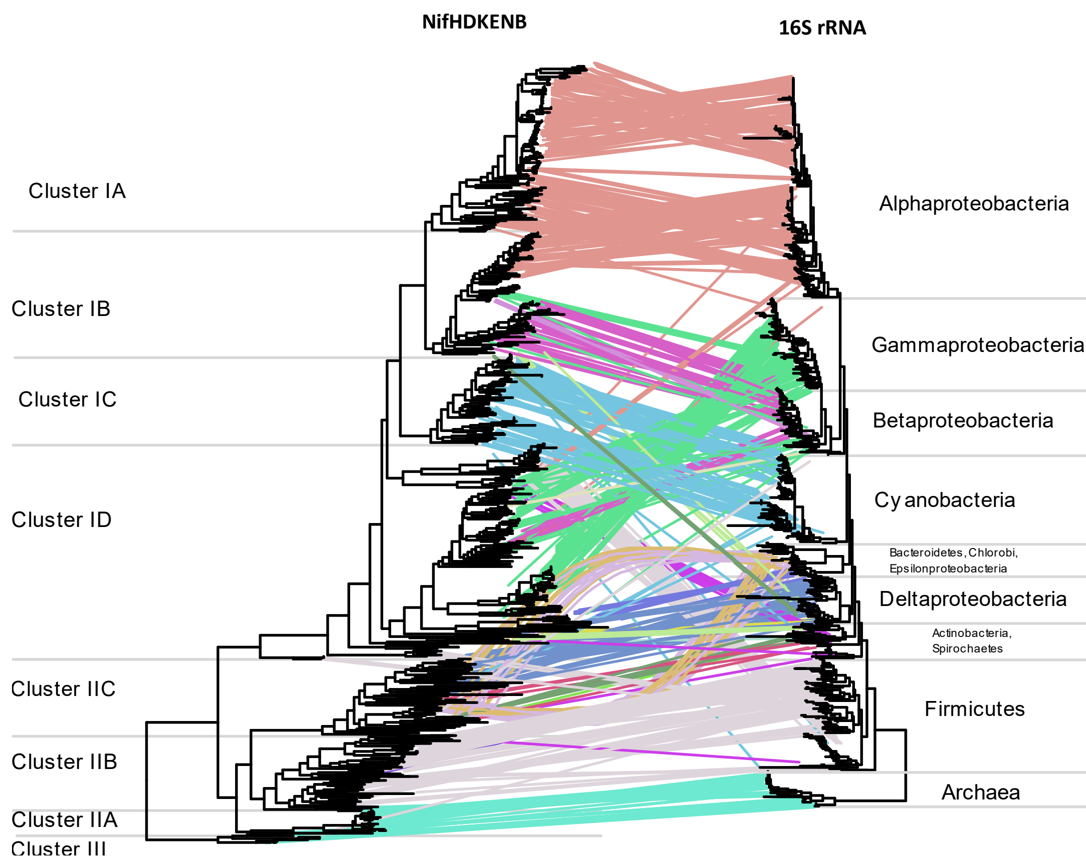


Figure 2. 9 Tanglegram comparing the concatenated NifHDKENB tree with 16S rRNA phylogeny of the diazotrophs, both obtained using FastTree. Lines indicate the respective positions of the 963 bacteria in the 2 trees. See Supplementary Tree file for details of the horizontal gene transfer.

2.4 Discussion

Using the minimal structural (NifHDK) and assembly (NifENB) protein components of the nitrogenase enzyme as a marker of diazotrophy, we found diazotrophs in the Proteobacteria, Firmicutes, Cyanobacteria, Bacteroidetes, Chlorobi, Nitrospira, PVC group, Spirochaetes, Deferribacteres, Aquificae, Fusobacteria, Chloroflexi, and Euryarchaeota. In addition to these phyla, diazotrophy has been reported in Chrysiogenales (Dos Santos et al. 2012), Acidobacteria (Kapili et al. 2020) and Elusimicrobia (Zheng et al. 2016), and *Azoarcus-Aromatoleum* groups (Raittz et al. 2021) but these were not included in this study as at least one gene was absent or misannotated

according to the KEGG database. Although the Candidate Phyla Radiation group (CPR) accounts for over 15% of bacterial diversity (Danczak et al. 2017), no evidence of diazotrophy was found in the genomes available for this prokaryotic group. CPR bacteria have reduced genome size and lack basic pathways like the citric acid cycle and respiratory chains (Hug et al. 2016), indicating the primitive nature of these bacteria. Among non-CPR bacteria, Proteobacteria accounted for more than half of the diazotrophic species and some smaller phyla are represented by only a single species. While distribution of diazotrophs appears widespread across the non-CPR bacteria, only 3.9% of the bacterial whole genomes harbor the six core *nif* genes (Figure 2.2a). With the consumption of 16 ATPs per dinitrogen reduced, the nitrogenase system is very costly for a bacterium; hence it is energy efficient for species living in nitrogen-rich environments to rescind the nitrogenase enzyme. In fact, multiple studies have shown a decrease in diversity of diazotrophs when treated with nitrogen fertilizer (Feng et al. 2018),48]. Phyla represented by photosynthetic species like Cyanobacteria (22.2%) and Chlorobi (41.2%), that are primarily found in nitrogen deficient aquatic habitats, have significantly higher representation of diazotrophic members. This indicates that the maintenance of nitrogenase systems is heavily dependent on the scarcity of nitrogen in the niche, and coupling of the two most important biological pathways, photosynthesis and nitrogen fixation, is a more efficient method in fixing both C and N.

Traditionally, phylogeny of nitrogenase has been based on either *nifH* or *nifD* but use of one or both of these was reportedly prone to a false positive rate of diazotrophy (Dos Santos et al. 2012). By using all six genes, we found 84% of the sequenced isolates to have biochemical evidence, and the remaining 16% are recently isolated strains (after

2010) (Supplementary Figure S11 https://www.mdpi.com/2076-2607/9/8/1662/s1). None of the information available for these strains included mention of experimental testing for diazotrophy, so these could all be new potential diazotrophs not yet characterized. Conversely, the literature abounds with reports of bacterial genera able to fix nitrogen but without genetic or biochemical evidence, for example the genus *Bacillus*. Most *Bacillus* found to have *nif* genes, such as *B. polymyxa* and *B. macerans*, have been shown to have multiple plant beneficial properties along with nitrogen fixation, and have been reclassified as *Paenibacillus* (Liu et al. 2019). *B. nealsonii* and B *caseinilyticus*, the only *Bacillus* that possesses all core six *nif* genes, are extremophiles and lack biochemical evidence for nitrogen fixation. This is in contrast to multiple claims of diazotrophic *Bacillus* isolates, supported only by culture evidence for growth on nitrogen free media (Çakmakçı et al. 2007a; Ding et al. 2005; Rilling et al. 2018; Yousuf et al. 2017). However, none of these claims have been supported by genetic evidence. In addition to the archaeal diazotrophs proposed by Leigh (Leigh 2000), potential diazotrophic species were observed in the orders *Methanocellales* and *Methanomicrobiales* as well. Although Actinobacteria is a large phylum where diazotrophy has been reported in multiple genera (Gtari et al. 2012), only *Frankia* and *Propionibacterium* were found to have all six *nif* genes in this study.

The six gene criterion yielded nitrogenases aligning exclusively with Clusters I, II and III in the Raymond classification (Raymond et al. 2004). While several genomes contained *nifH* aligning with Clusters IV or V, none had all six genes, with at least one of the *NifENB* gene homologs either missing or not annotated in the KEGG database. As *nifH* of Clusters IV and V do not encode functional nitrogenases, this result strengthens

the case for the use of six genes for probing potential diazotrophs. One notable exception is *Endomicrobium proavitum* reported by Zheng et al. (Zheng et al. 2016) which was experimentally shown to fix nitrogen but harbors nitrogenase homologous to Group IV nitrogenase. However, *Endomicrobium* genomes available in the KEGG database lack *NifEN* genes. Many operons containing alternative nitrogenases *vnfHDK* or *anfHDK* did not have their own set of *ENB* genes. This is perhaps indicative that some bacteria encoding Group IV—NifHDK can express an active nitrogenase, even if they do not encode the NifEN assembly proteins. The exact mechanism for this is yet to be determined but suggests that there may be some exceptions to the six gene criteria established by Dos Santos et al. in 2012. Phylogenetically, the six *nif* genes have very similar distribution, suggesting that the *nif* operon evolved as a unit. Our analysis of the six genes points to the influence of habitat on the evolution of the nitrogenase enzyme. Smaller phyla with species adapted to a particular niche were found clustered together in a single cluster. Taxonomically unrelated species occurring in a particular niche were found to have similar *nif* genes, suggesting the role of lateral gene transfer in evolution of the nitrogenase enzyme. As an example, two distinct physiological groups of Firmicutes fell in two distinct NifHDKENB clusters. Anaerobic clostridia and related taxa aligned with archaea occurring in Cluster IIB, while the aerobic firmicutes occurred in Cluster I together with the Actinobacteria. On the other hand, Cyanobacteria, which comprise photosynthetic and mostly aquatic bacteria, occurred in a monophyletic clade which is in close agreement with its phylogeny supported by the study of Esteves-Ferreira et al. (Esteves-Ferreira et al. 2017). Species of larger phyla like Proteobacteria and Firmicutes which are adapted to a wide range of physiological conditions were found to have *nif*

genes distributed across multiple clusters in the phylogenetic tree (Figure 2.4). Proteobacteria form phylogenetically coherent clusters except for *Bradyrhizobiaceae* and related families of Alphaproteobacteria (Figure 2.5), and Betaproteobacteria cluster together with acidophilic, methanotrophic bacteria from diverse taxonomical clades also observed by Khadem et al. (Khadem et al. 2010) with methylotrophic Verrucomicrobium, *Methylacidiphilum.* Subcluster ID mainly consists of Gammaproteobacteria with some exceptions of Alphaproteobacteria like *Martellela*, *Cohesibacter*, *Magnetovibrio*, and Betaproteobacteria like *Aquaspirillium*, *Sideroxydans*, and some *Rhodocyclales.* Similarly, nitrogenase in Archaea is confined within the methanogenic Euryarchaeota. The Stenosarchaea group of Euryachaeota comprises most of the Archaeal diazotrophs, all of which occur in Cluster II together with other anaerobic bacteria. Smaller phyla like Bacteroidetes, Chlorobi, Spirochaetes, and the PVC group, which, in spite of the wide range of habitats, all clustered together in subcluster IIC, correlating with their anaerobic respiratory pathways.

Raymond et al. (Raymond et al. 2004) proposed two plausible hypotheses of the evolution of nitrogenase. The first is that the nitrogen-fixing LUCA harbored nitrogenase, and present-day diversity was attained by loss of *nif* genes in large numbers of taxonomic groups including Eukaryotes and non-methanogenic archaea. The second hypothesis has nitrogenase originate in the methanogenic archaea, with transfer of *nif* genes to bacteria by HGT. Our results strongly support the latter hypothesis. The absence of *nif* genes in Candidate Phylum Radiation (CPR), Eukarya, and non-methanogenic Archaea, and branching of bacterial *nif* from archaea strongly suggests that nitrogenase originated first in primitive methanogenic Archaea. This hypothesis has been supported by Boyd and

Peters (Boyd and Peters 2013). The present phylogenetic analysis of *nifHDKENB* genes and the chronogram of prokaryotes proposed by Zhu et al. (Q. Zhu et al. 2019) (Supplementary Figure S10 https://www.mdpi.com/2076-2607/9/8/1662/s1) enabled us to propose the evolution of nitrogenase in each phylum harboring diazotrophic species (Figure 2.10). Kasting and Walker (Kasting and Walker 1981) proposed that a nitrogen crisis occurred at ~3.5 Ga, overlapping with the origin of methanogenic Euryarchaeota (Supplementary Figure S10 https://www.mdpi.com/2076-2607/9/8/1662/s1); hence nitrogenase must have evolved during that time in ancient methanogenic Euryarchaeota in response to the shortage of combined nitrogen. Ancient nitrogenase was vertically transferred to two groups of Euryarchaeota, Methanomada, and Stenosarchaea. An alternative nitrogenase originated from ancient Methanomada and forms Cluster III. The first *nif* in bacteria was laterally transferred to ancient bacteria from the ancient Methanomicrobia between 3.5 and 3.18 Ga, after which it was vertically transferred to all bacteria (Figure 2.10). The present scattered distribution of nitrogenase across multiple phyla likely originated through loss of nitrogenase by organisms adapted to live in nitrogen-rich environments, followed by multiple lateral transfers within anaerobic niches. The great oxidation event in 2.3 Ga. must have turned many bacterial environments aerobic, creating isolated anaerobic niches. This would have brought diverse anaerobes into close proximity in the available anaerobic niches. Co-occurrence of unrelated taxa in close proximity would increase exchange of genes, thus favoring HGT among anaerobes. This is supported by the similarity of *nif* genes of anaerobes like Bacteroidetes, Chlorobi, Spirochaetes, Deltaproteobacteria, Clostridium, and PVC to Methanomicrobia. Based on the branching pattern in Cluster II, two distinct HGT’s must

have occurred to achieve the present distribution of *nif* genes among anaerobic bacteria (Figure 2.10), where Clostridium II and Fusobacteria obtained the genes much later than other anaerobes. Cluster I on the other hand is dominated by aerobic and facultative anaerobic diazotrophs, most of which belong to Alphaproteobacteria, Cyanobacteria, and Firmicutes. Although the majority of Alphaproteobacteria have lost the *nif* genes, they are preserved in the species which evolved to be in symbiosis with plants (*Rhizobiales*) and free-living species which primarily occur in nitrogen deficient habitats like water (*Rodospirillales* and *Rhodobacterales*). Occurrence of Aquificae and Nitrospirae at the root of Cluster I and clustering of Cyanobacteria, Actinobacteria, and aerobic firmicutes in the *nif* tree strongly suggests nitrogenase in these phyla must have evolved by vertical transfer from ancient non-CPR bacteria.

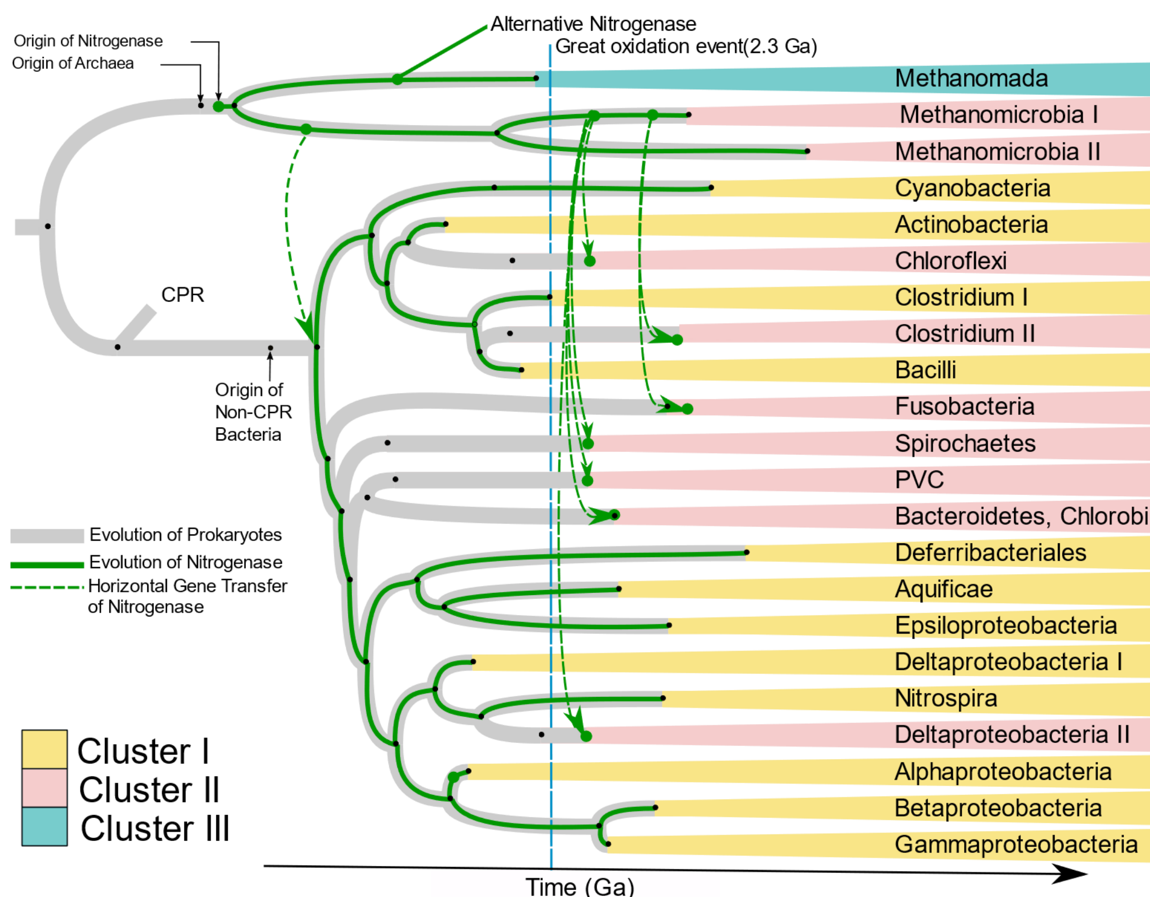


Figure 2. 10 Proposed evolution of Diazotrophs. The gray tree represents the chronogram of prokaryotes while the green tree represents the proposed evolution of nitrogenase. The dashed green lines indicate horizontal gene transfer as supported by homology of NifHDKENB sequences. Black dots indicate origin of the respective phyla or groupings. The evolutionary timescale is based on Zhu et al., 2019 (Q. Zhu et al. 2019). The Methanomada clade in Euryarchaeota includes orders Methanobacteriales and Methanococcales. Methanomicrobia I includes the orders Methanocellales and Methanomicrobiales. Methanomicrobia II includes Methanosarcinales. Clostridium I includes the family Peptococcaceae, Clostridium II includes Clostridia and Negativicutes. Deltaproteobacteria I includes Myxococcales, and Deltaproteobacteria II includes other Deltaproteobacteria.

2.5 Conclusion

This study shows that the structural genes of nitrogenase (*nifHDK*) and Mo-Fe cofactor assembly proteins (*nifENB*) have very similar phylogeny, indicating that all six genes must have co-evolved. Additionally, the similarity of NifE with NifD, and NifK with NifN strongly supports the origin of NifEN by duplication of an ancestor NifDK.

Using these six genes as criteria for the study of diazotophs, only potentially active diazotrophs were selected. All nitrogenases selected using this criterion could be classified into one of the Raymond Clusters, I, II, or III. Using all six genes also enabled us to understand the diversity of nitrogenase in individual phyla and the relationship between the nitrogenases in different phyla. Lateral gene transfer has played an important role in distribution of nitrogenase, hence multiple taxa in the same phylum can have very distinct nitrogenases based on their physiological and ecologically driven co-occurrence with unrelated taxa. Therefore, this work provides reference for any inter-phyla comparison of *nif* sequences and a quality database of Nif proteins that can be used for identifying new *nif* sequences and designing *nif* primers.

Chapter 3

Study of Diversity of Free-living Nitrogen Fixing Bacteria in Sioux Prairie by Culture-based Approach.

The work reported in this chapter was conducted by Amrit Koirala in collaboration with Nabilah Ali Alshibli.

3.1 Introduction

Nitrogen is the fourth most abundant component of living organism, from bacteria to animals. It is required for the synthesis of many biomolecules like amino acids, vitamins, co-factors, nucleotides, and porphyrin (e.g., chlorophyll). It is also the most important factor limiting plant yields in terrestrial and aquatic ecosystems (Zehr et al. 2003). The major reservoir of nitrogen is atmosphere which contains around 78.1% N_2 but it needs to be reduced to ammonia before majority of living organisms can assimilate it into their biomass (Belnap 2002). Plants obtain these combined forms of nitrogen by natural or more recently synthetic processes. The Haber-Bosch process is the most used artificial process used to reduce ammonia from N_2 for use as fertilizer. Natural processes involve biological nitrogen fixation and to a minor degree lightning (Vance 2001).

Biological nitrogen fixation (BNF) is carried out by a diverse group of prokaryotes from Archaea and Bacteria, termed diazotrophs. Diazotrophs are divided in two major types: 1) symbiotic nitrogen – fixing bacteria which form a symbiotic relationship with legumes like *Rhizobium*, with actinorhizal plants such as *Frankia*, and *Cyanobacteria* associated with cycads, and 2) free-living nitrogen fixers belonging to genera such *as Burkholderia, Azotobacter, Azospirillum,* and *Clostridium* (Postgate et al.

1982). Although the nitrogenase enzyme complex is the only enzyme system in prokaryotes known to fix nitrogen, some studies have indicated the existence of alternative enzymes which can reduce the dinitrogen present in the atmosphere into a combined form like ammonia. Higdon et al. in 2020 used genomics and a machine learning model to identify putative BNF components even in absence of *nif* genes (Higdon, Huang, et al. 2020).

Biological nitrogen fixation (BNF) is the primary (in pre-industrial age) and the second largest contributor (in post-industrial age) of fixed nitrogen to the biosphere (Vitousek et al. 2013). It is estimated to contribute 128 Tg N yr^{-1} in terrestrial ecosystem and 140 Tg N yr^{-1} in marine ecosystem (Fowler et al. 2013). Most of the marine diazotrophs (Zehr and Capone 2020), terrestrial cyanobacteria (Esteves-Ferreira et al. 2017), methanogenic Archaea (Adam et al. 2017), and Firmicutes (Achouak et al. 1999) are able to fix nitrogen on their own. Since nitrogenase is irreversibly inactivated by O_2 and most of these diazotrophs growing in aerobic condition, free-living diazotrophs need to spend additional cellular resource to decrease redox potential around nitrogenase. Hence, free-living diazotrophs are less efficient and fix about one-tenth of the total atmospheric N_2 fixed by symbiotic diazotrophs (Unkovich and Baldock 2008). However, given the omnipresence of these bacteria in virtually any habitat, nitrogen fixation by free-living diazotrophs has been predicted to contribute a significant amount of fixed nitrogen towards biological nitrogen fixation (Vitousek et al. 2013). Some diazotrophs belonging to the genera *Paenibacillus, Herbaspirillium, Klebsiella,* and *Azospirillum* occur in close association with plants in the rhizosphere. They are termed associative diazotrophs and depend on plant exudates for carbon source (Van Dommelen

and Vanderleyden 2007). For the purpose of this research free-living nitrogen fixing bacteria also include the associative diazotrophs as it is very difficult to distinguish between associative and free-living bacteria in the soil environment heavily occupied by sod forming grasses.

Dinitrogen is a complex hetero-enzyme made up of two components. Component I is also called nitrogenase reductase or Fe- protein, is a homodimer of NifH and coded by *nifH.* The second component is dinitrogenase and is a hetero-tetramer consisting of two proteins. The alpha-chain is coded by *nifD* and the beta-chain by *nifK* (Hu and Ribbe 2015). Of these three structural components *nifH* genes have been used extensively for the study of diversity of the diazotrophs (Zehr et al. 2003) and multiple universal and group specific primers has been developed for the study of *nifH* gene (Gaby and Buckley 2012a). Phylogeny based on *nifH* gene classifies *nifH* and its homologs into the following clusters: Cluster I consist of NifH and VnfH from aerobic and facultative anaerobic proteobacteria and cyanobacteria. Cluster II consist of VnfH and NifH from some archaea. Cluster III consist of NifH from strict anaerobes and Cluster IV includes all NifH homologs like NflH and Bch/ChlL (Zehr et al. 2003).

All the *nifH* primers developed so far are degenerate and none of the universal primers developed have 100% coverage of the diazotrophic community (Poly, Monrozier, and Bally 2001; Ueda et al. 1995). In addition to this, *nifH* primers are prone to amplifying homologous genes from cluster IV which do not have any role in nitrogen fixation. Therefore, a culture-independent method using *nifH* as marker gene is prone to bias caused by primary degeneracy and low coverage. In addition to this, a culture

independent study fails to provide the bacterial cultures which can be further characterized. Therefore, several culture based methods have been developed for the study of diversity of diazotrophs (Dobereiner et al. 1976; Mirza and Rodrigues 2012; Park et al. 2005). As nitrogenase is extremely sensitive to oxygen most of these methods relied on use of semisolid agar or broth to limit the oxygen exposure, but use of solid media in reduced oxygen environment has been found to increase the diversity of culturable potential diazotrophic bacteria (Mirza and Rodrigues 2012).

Prairies are vast extents of arid/semiarid grassland, occupying the vast majority of the central United States (Weaver 1954). The biological soil crust (BSC) in these grassland presents a unique ecosystem where diverse microorganisms live in close proximity with grass roots and dead and decaying organic matter (Risser 1985). BSCs are the major source of nitrogen in arid and semiarid lands (Belnap 2002) . In areas with minimum human activities, microorganisms present in BSC play a vital role in fixation of carbon and nitrogen (Steven et al. 2015). Free-living diazotrophic bacteria contribute to a substantial amount of nitrogen in grassland ecosystems where symbiotic nitrogen fixation is rare (Taylor et al. 2019). A study of asymbiotic nitrogen fixing microorganisms in natural, sown, and partially degraded alpine grassland on the Tibetan Plateau identified several diazotrophic bacteria from phylum Proteobacteria and observed significant differences in diversity of asymbiotic nitrogen fixing bacteria in these grasslands (Li et al. 2021). Similarly, a study of nitrogen fixing bacteria associated with Switchgrass in native tallgrass prairie of north Oklahoma reported several species of Alphaproteobacteria and Firmicutes having *nifH* genes. Although other studies (Davis et al. 2010; Patra et al. 2006) have also shown that nitrogen fixation in soil is significantly affected by the grass

species abundant in the given grassland, the physiology and limitations of free-living diazotrophs in natural grassland ecosystems are virtually unexplored, signifying the need for further study (Vitousek et al. 2013). Therefore, the purpose of this study was to explore the diversity of free-living nitrogen fixing bacteria in natural prairie and use the isolates obtained to characterize the nitrogen fixation in this grassland ecosystem.

3.2 Materials and methods

3.2.1 Sampling site

Soil samples were collected from Sioux Prairie, which is a natural prairie grassland preserved and maintained by The Natural Conservancy (TNC). It is located at 231st Street, Colman, SD 57017 (44°2’8” N 96°46’0” W) (Figure 3.1) at 1720 ft elevation. This site has typical mid-west, continental climate with four distinct seasons, where winters are cold and dry, and summers are warm and semi-humid. The average annual precipitation is 635 mm and the highest temperature in summer is close to 90°F and lowest in winter on average is 10°F.

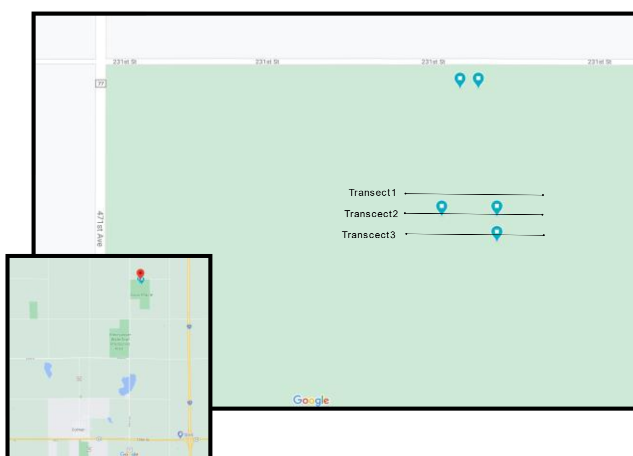


Figure 3. 1 The location of sampling sites at The Nature Conservancy (TNC) site at Sioux Prairie, South Dakota. Samples were collected from three parallel transects which were 5 meters apart. Source: google map

Samples were collected from three transects located 5 meters apart and 9 samples were collected from each transect resulting in a total of 27 samples. In each transect,

sampling sites were located 1 meter apart. Soil cores were collected to 10 cm below the surface and were 2.5 cm in diameter. All the soil samples were collected in airtight, sterile vials and stored at 4°C when transferring to the laboratory for further processing.

3.2.2 Isolation of potential diazotrophic bacteria

Potential diazotrophic bacteria were isolated using nitrogen free media (NFM) which supports the growth of nitrogen fixing bacteria only. For the primary isolation NFM solidified with Noble agar, Nitrogen free Agar (NFA) medium was used. NFA media comprises of sugars (Glucose (2g/l), Arabinose (2g/l), Mannitol(2g/l), Malic acid (2g/l)) and K_2HPO_4 (0.2g/l), KH_2PO_4 (0.5g/l), $MgSO_4.7H_2O$ (0.2g/l), $FeSO_4.7H_2O$ (0.1g/l), $Na_2MoO_4.2H_2O$ (0.005g/l), NaCl (0.2g/l) and solidified with noble agar (15 g/l, Difco, Catlog No. 214230). NFM broth used for purification in later stages has same composition without noble agar. The pH of NFM was 7.2.

The nitrogen free medium and the major constituents were tested for organic and inorganic nitrogen compounds (NO_2^-, NO_3^-, NH_4^+, Total Nitrogen) at CAChE Nutrient Analysis Core Facility at Florida International University.

Soil samples (1 gm) were suspended in 10 ml sterile deionized water and mixed by shaking on a shaker for 8 h and serial dilutions were plated on nitrogen free agar (NFA) medium in replicates of six (Figure 3.2). Inoculated plates were incubated at 28 °C for 15 days in reduced oxygen condition (n=3) and ambient oxygen condition (n=3). Reduced oxygen condition was maintained by incubating in an airtight chamber with a gaspak system (BD GasPak™ EZ). After incubation, the number of colonies were

counted, and 8 isolates were selected from each sample from each incubation condition based on morphological variation. This yielded 432 isolates for further processing.

3.2.3 Purification of isolates to obtain pure cultures of potential diazotrophs

The isolates obtained were passed through a rigorous protocol to obtain pure cultures and to confirm their growth in both NFA and NFM broth (Figure 3.2). Isolates were streaked on NFA for 3 generations to eliminate the possibility of any carryover of nitrogen from the original sample supporting the growth of isolates. Isolates were then sub-cultured on nitrogen rich R2A (Difco™ R2A Agar) plates and incubated at 28 °C for a week in both aerobic and reduced oxygen condition. In cases where, the R2A revealed cocultures of two or more bacteria

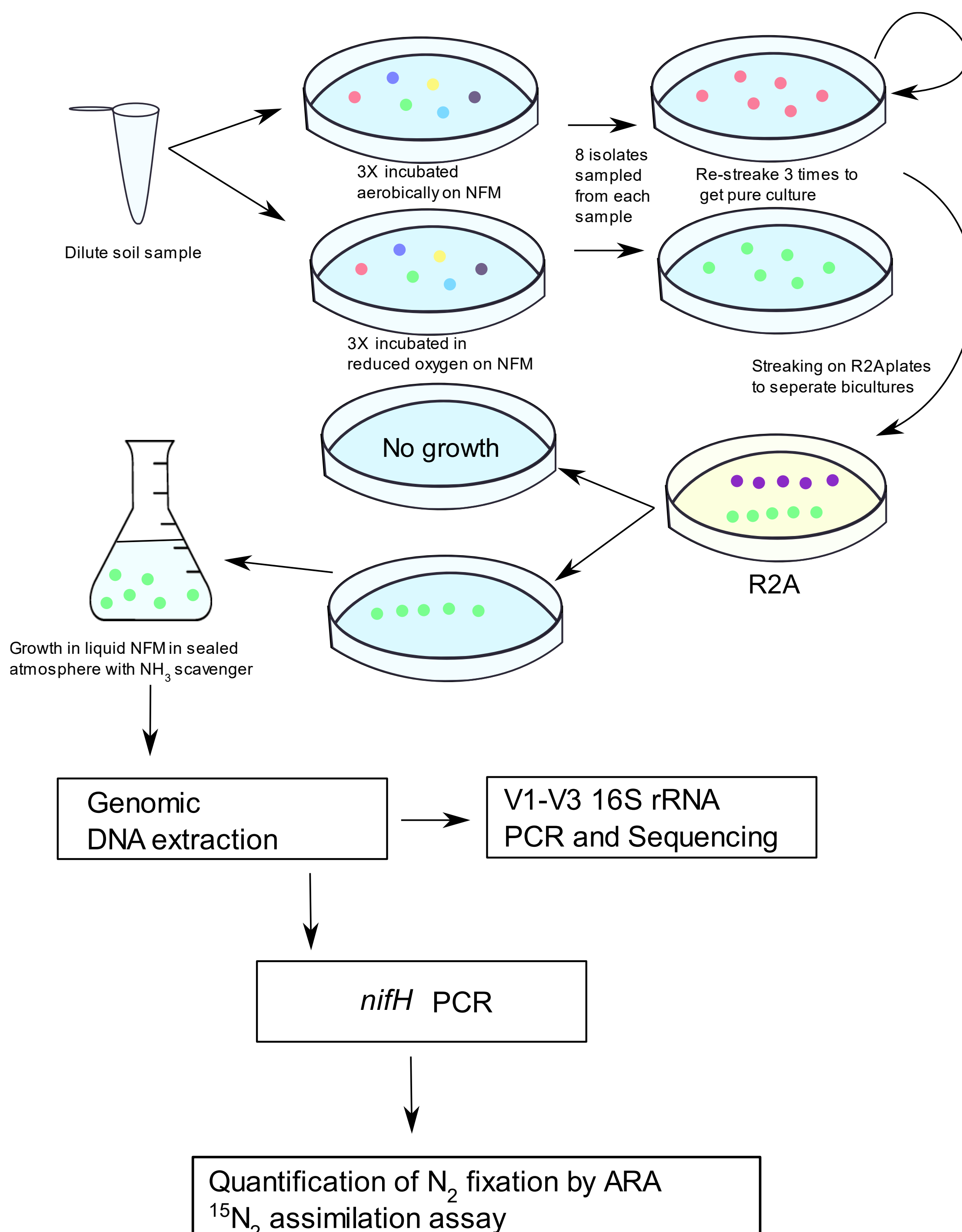


Figure 3. 2 Flow diagram illustrating isolation and processing of putative nitrogen fixing bacteria from Sioux Prairie soil samples.

of different colony morphology were obtained, each distinct colony was re-streaked on NFA to confirm their ability to grow on NFA by itself. Several isolates failed to grow on NFA and were discarded as non-diazotrophic co-cultures.

In some cases, the isolates were contaminated with fungal species, and were sub-cultured in R2A containing cycloheximide (200µg/ml) and nystatin (40 µg/ml). The absence of fungus in these isolates was later confirmed by PCR amplification of the eukaryotic internal transcribed spacer region (ITS) using primers ITS5 (5′-GGAAGTAAAAGTCGTAACAAGG-3′) and ITS4 (5′-TCCTCCGCTTATTGATATGC-3′) (Bruns, Fogel, and Taylor 1990).

The isolates that continued to grow on NFA, were sub-cultured on NFM broth and incubated in an airtight chamber with Clinoptilolite to remove any atmospheric nitrogen that could be salvaged, to filter out the isolates scavenging NH_3 from atmosphere (Tosun 2012). The isolates satisfying all the criteria were processed further as well as stored at -80 ^{o}C in 50% glycerol.

3.2.4 Extraction of genomic DNA

Total genomic DNA was extracted from each isolate growing in R2A broth for 48 h using the Microbial DNA extraction kit from MoBio. The amount of DNA obtained was quantified spectrophotometrically using NanoDrop. Quality was accessed electrophoretically by visualizing total genomic DNA resolved in a 1% agarose gel and stored at -20 ^{o}C.

3.2.5 Amplification of 16S rRNA sequence and classification of isolates

The hypervariable V1 to V3 region of the 16S rRNA gene was amplified using universal PCR primers 27F (AGA GTT TGA TCM TGG CTC AG) (Weisburg et al. 1991) and 518 R (GTA TTA CCG CGG CTG CTG G) (G Muyzer, de Waal, and Uitterlinden 1993) . The PCR was performed in 30µl reactions consisting of 0.2µl (5000U/ml) Taq polymerase (NewEngland BioLabs), 3µl (10X) PCR Buffer, 0.6µl dNTP (each 10mM), 0.6µl (10mM) each Primer, and 2.4µl (25mM) $MgCl_2$. The PCR conditions were initial denaturation at 95 ºC for 4 min, 30 cycles at 95 ºC for 30 s, 50 ºC for 45 s, and 72 ºC for 1 min, and a final elongation at 72 ºC for 10 min. The integrity and quality of PCR products were checked electrophoretically using a1% agarose gel, and sequenced using Sanger sequencing (GenScript Sequencing company, Piscataway, NJ). The sequences were aligned and classified using the Sina Sequence Aligner against the SILVA database (https://www.arb-silva.de/aligner/). The aligned reads with references were used to calculate the maximum likelihood tree using PhyML, using a Bayesian-like transformation of aLRT (aBayes) bootstrapping (Guindon et al. 2010). The phylogenetic tree was visualized and annotated using iTOL (Letunic and Bork 2019).

3.2.6 Amplification and classification of *nifH* sequences.

Amplification of *nifH* was performed using multiple primer pairs (Table 3.1) at different reaction condition. The PCR was performed in 30µl reactions comprising of 0.2µl (5000U/ml) Taq polymerase (NewEngland BioLabs), 3µl (10X) PCR Buffer, 0.9µl dNTP (each 10mM), 0.9µl (10mM) each primer, and 2.5µl (25mM) $MgCl_2$. The PCR

condition for amplification of PolF/PolR were initial denaturation at 94 °C for 5 min followed by 30 cycles of denaturation at 94 °C for 1 min, annealing at 55 °C for 1 min, and elongation at 72 °C for 30 seconds with final elongation at 72 °C for 5 minutes. The product size for PolF/PolR PCR is ~360 bp. To optimize PCR conditions nonspecific binders like BSA (NewEngland BioLabs), 5% DMSO, and 5% Glycerol were tested. For IGK3/DVV and Ueda19F/Ueda407R primer pairs annealing temperatures were 59°C and 52°C. The product size with these primers ranges from ~360 to 400 bp.

Table 3. 1 Primer pairs used for amplification of nifH genes in this study

Primers	**Reference**
PolF TGC GAY CCS AAR GCB GAC TC PolR ATS GCC ATC ATY TCR CCG GA	(Poly et al. 2001)
IGK3 GCI WTH TAY GGI AAR GGI GGI ATH GGI DVV ATI GCR AAI CCI CCR CAI ACI ACR TC	(Gaby and Buckley 2012a)
Ueda19F GCI WTY TAY GGI AAR GGI GG Ueda407R AAI CCR CCR CAI ACI ACR TC	(Ueda et al. 1995)

The PCR products were subjected to quality check on 1% agarose. For cases with multiple bands, the gel portion containing the band of appropriate size was cut and extracted using Zymogen gel cleaning kit (Zymoclean Gel DNA Recovery Kit) and PCR was run again. The PCR products that passed the quality test were sequenced. The sequences were aligned against uniprot protein database (https://www.uniprot.org/blast/) using NCBI BLAST to filter out sequences from non-specific amplification. Only those

sequences that aligned to a NifH protein in the uniprot database were used for further analysis. The phylogenetic assignment of the *nifH* sequences was done by aligning them with the reference *nifH* sequences from my previous study (Chapter 2).

3.2.7 Quantification of nitrogen fixation by selected isolates

Isolates were selected based on their growth rate and relative abundance for further characterization. Acetylene Reduction Assay (ARA) and ^{15}N assimilation assay was employed for indirect and direct measurement of nitrogen fixation. Cultures were grown in airtight serum vials (60 ml) with 1% headspace replaced by acetylene gas for ARA. The headspace was analyzed for the presence of ethylene using gas chromatography available at Functional Genomics core facility at SDSU.

The ^{15}N assimilation assay was performed by growing isolates in airtight serum vials with 10% head space enrichment with $^{15}N_2$ (Cambridge Isotopes) for 14 d. The total biomass was harvested by centrifugation (10,000Xg for 5 min) and dried. ^{15}N assimilation was quantified by mass spectrometry at the Stable Isotope Facility at UC, Davis, California. For both quantification methods E. *coli K*12 and *H seropedicae* ATCC 35892 were used as negative and positive controls respectively.

3.3 Results

3.3.1 Chemical analysis of NFM

A very small amount of nitrogen was detected in the laboratory NFM and its constituents and some of them were below the laboratory method detection limit (Table 3.2). The deionized water and NFM media prepared from it had 4.70 and 11.53 µmol/L

total nitrogen. The Millipore pure water and the NFM prepared from it had only 2.57 and 6.42 µmol/L of total nitrogen. All the measurements of total nitrogen were below the laboratory practical quantitation limit (PQL). The sum of nitrate and nitrite (N+N) was also either below or close to the practical quantification limit (0.96 µmol/L). There was not any detectable nitrogen in the agar and sugar used for preparing these media.

Table 3. 2 Chemical analysis of NFM, different waters available in lab, arabinose and Noble agar used for solid NFM.

Sample	**NO_3-N (µmol/L)**	**NO_2-N (µmol/L)**	**NH_4-N (µmol/L)**	**TN (µmol/L)**	**N %**
Deionized water	0.24	**0.04**	**0.01**	**4.70**	
Millipore pure water	0.16	**0.01**	**0.16**	**2.57**	
NFM using distilled water	0.89	1.14	2.26	**11.53**	
NFM using Millipore pure water	0.97	0.31	1.19	**6.42**	
Noble agar					ND
Arabinose					ND

Values in green are below the Laboratory Practical Quantitation Limit (PQL) and greater than or equal to Laboratory Method Detection Limit (MDL).

Values in red are less than MDL, ND = not detected.

Ammonia is the first choice as source of nitrogen for many bacteria and the amounts of nitrogen in the form of ammonia in these samples were also either below or close to PQL (1 µmol/L). The NFM prepared from deionized water had 2.26 µmol/L and NFM prepared from Millipore pure water had 1.19 µmol/L. Both concentrations are very low to support the growth of bacteria unless they can fix nitrogen from the atmosphere.

3.3.2 Enumeration and colony characteristics of putative diazotrophs

Bacteria with diverse morphological features were isolated from all 27 samples in both incubation conditions. Isolates on NFA produced colonies of less than 1 mm in size and were mucoid or dry and cream or white colored. On subsequent incubation for more than 14 d most of them produced a wrinkled surface, indicating formation of spores (Figure 3.3).

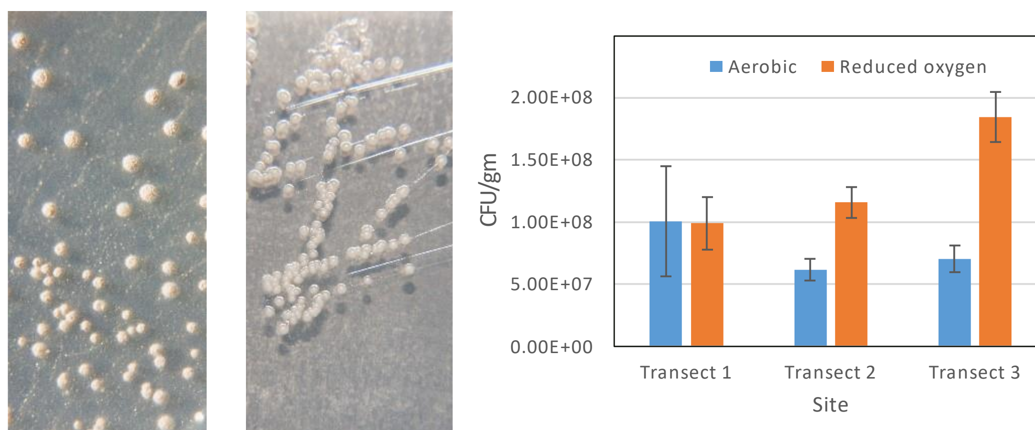


Figure 3. 3 Colony characteristic of a typical isolate on R2A and NFA (left) and culturable count of putative diazotrophs incubated under aerobic and reduced oxygen condition at 28°C for 14 d (right).

The culturable count was found highest in transect 3 and under reduced oxygen condition (Figure 3.3), however no significant difference of CFU/ml was found either across transect or between aerobic or reduced oxygen incubation conditions (p-value = 0.125).

3.3.3 Diversity of potential diazotrophs

Eight isolates were sampled from each of the 27 samples from plates incubated in both aerobic and reduced oxygen condition yielding 432 isolates in total. These isolates were further processed to obtaining single cultures, requires multiple subcultures on

Nitrogen free Agar (NFA) and transition between nitrogen rich R2A and nitrogen deficit NFA. When sub-cultured on R2A, most of the isolates produced distinct colony morphologies and also two or even three isolates. Both the isolates obtained on R2A were further screened on NFA and usually one of the two isolates failed to grow on NFA, indicating the co-existence of non-diazotrophic bacteria in co-culture with diazotrophic bacteria. The non-diazotrophic bacteria were discarded and not pursued further. Some of the isolates yielded fungi together with bacteria, and these which were eliminated by adding cycloheximide and Nystatin. The isolates satisfying our criteria were then evaluated for their ability to grow in liquid NFM without agar.

This rigorous process of purification yielded 486 putative diazotrophic bacteria in pure culture. The V1-V3 16S rRNA gene sequences were checked against RDP for their sequence quality, which yielded 474 partial 16S rRNA sequences that produced a significant match. These 474 sequences were classified using the SILVA database at 80% cutoff parameter. Of these 93% of the partial 16S rRNA sequences (442) were classified as Bacteria and remainder had a similarity percentage of less than 80% so labelled as unclassified. All the isolates belonged to one of the four phyla: Actinobacteria (198, 42%), Proteobacteria (207, 43%), Firmicutes (28, 6%), and Bacteroidetes (9, 2%) (Figure

3.3). Proteobacterial isolates fell into Alphaproteobacteria (87, 18%), Gammaproteobacteria (91, 19%) and Betaproteobacteria (29, 6%) (Figure 3.4a).

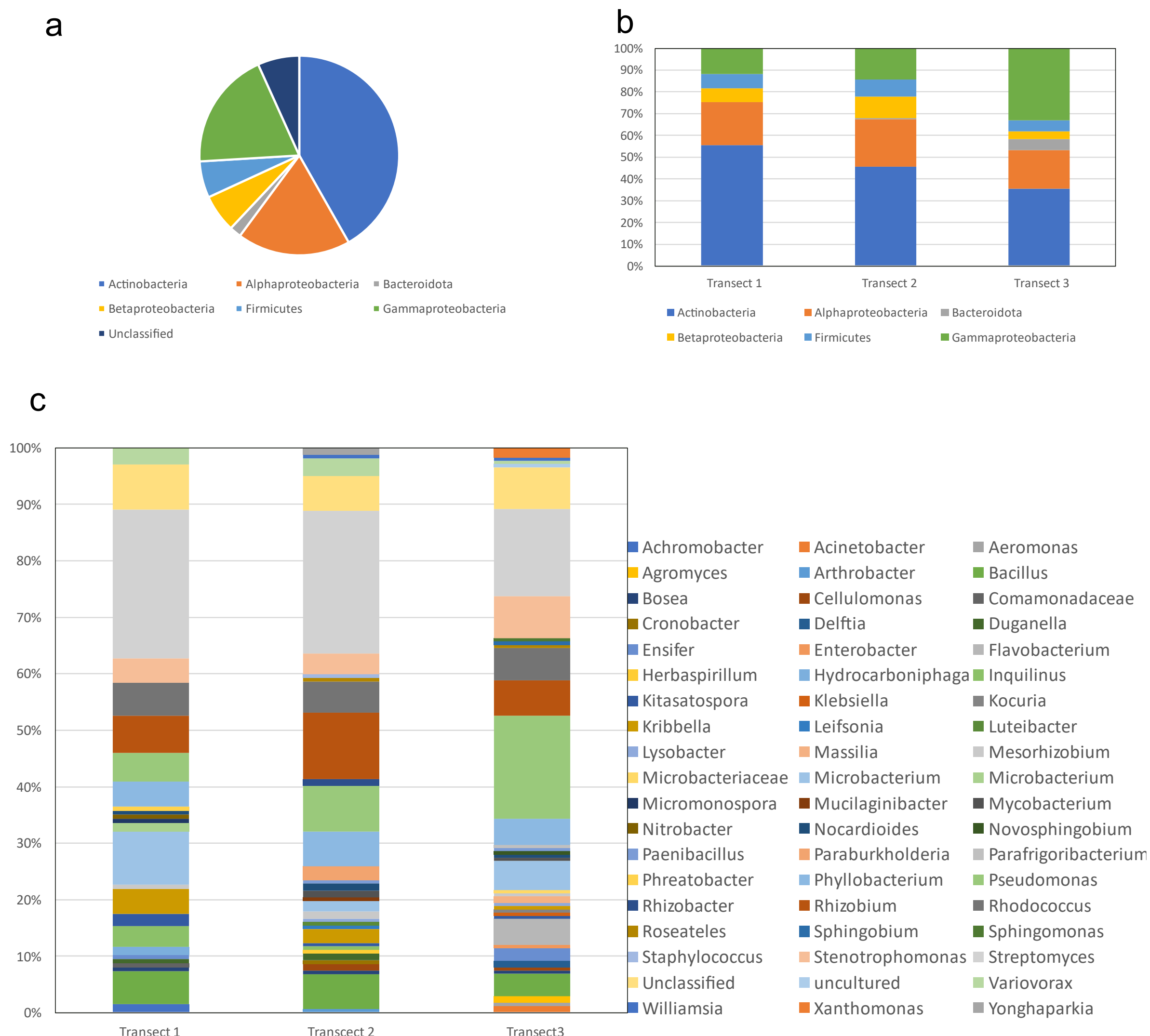


Figure 3. 4 Diversity of potential diazotrophic bacteria by V1-V3 region of 16S rRNA sequences. Distribution of total isolates by phylum (a), distribution of phylum across three transect (b), and distribution of genera in the three different transects (c).

<u>Distribution of isolates by site</u>

The isolates were quite uniformly distributed across the three different transects with some variations (Figure 3.4 b, c). Actinobacteria were dominant in transect 1 but Gammaproteobacteria were dominant in transect 3. Also, Bacteroidetes were absent in

transect 2. Of the 57 genera detected, *Bacillus, Bosea, Kitasatospora, Kribbella, Mesorhizobium, Microbacterium, Nocardioides, Phyllobacterium, Pseudomonas, Rhizobium, Rhodococcu, Stenotrophomonas, Streptomyces,* and *Variovorax* were present in all three transects.

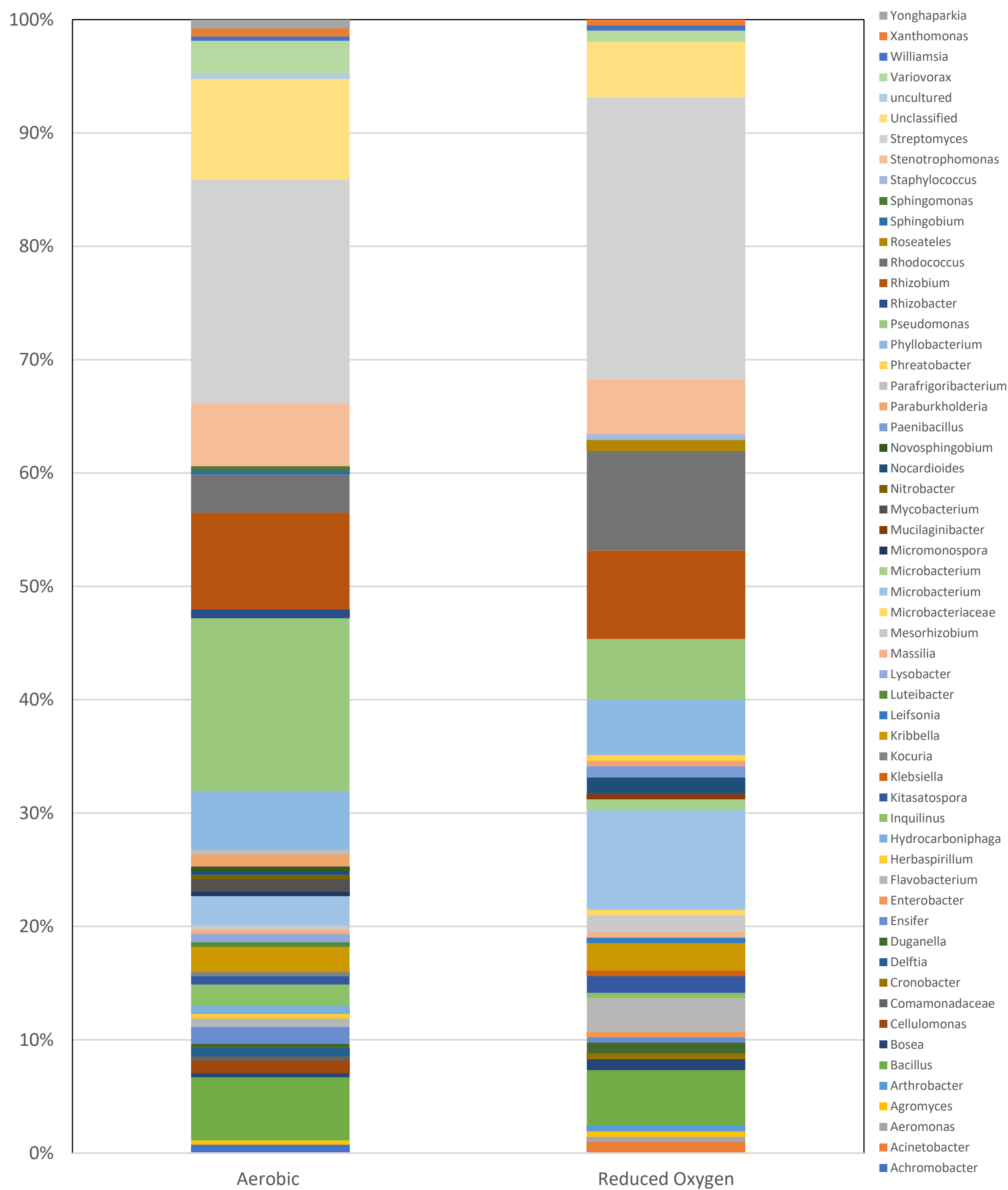


Figure 3. 5 Distribution of potential diazotrophs by incubation condition.

Distribution by incubation condition

Of the 57 genera, 24 were present in both incubation conditions, most abundant of these being *Streptomyces, Pseudomonas, Rhizobium, Rhodococcus, Stenotrophomonas, Phyllobacterium, Microbacterium,* and *Bacillus* (Figure 3.5). Nineteen genera were present in the aerobic incubation condition only and 14 genera were unique to reduced oxygen conditions (Appendix 1).

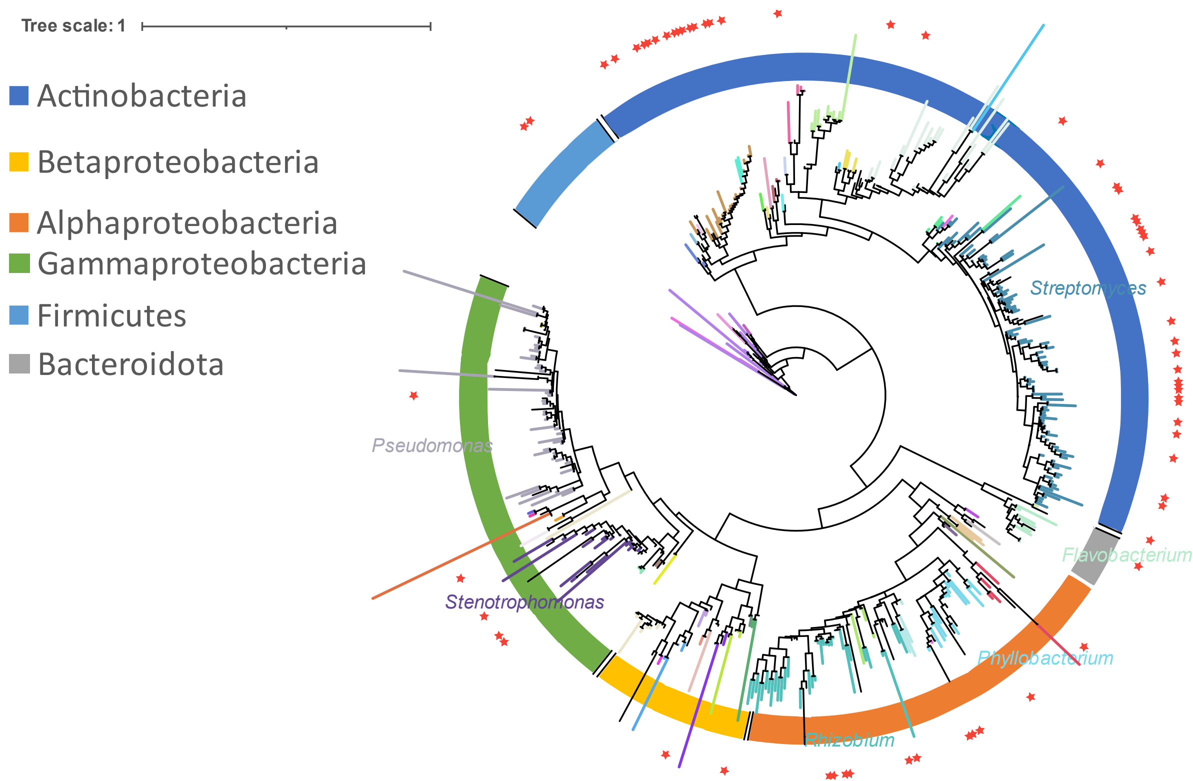


Figure 3. 6 Molecular phylogenetic analysis of partial 16S rRNA sequences obtained from potential diazotrophic isolates obtained by maximum likelihood in PhyML using aBayes analysis method. Branch tips are colored according to the genus classification and unclassified sequences not presented. The red stars on the outer ring indicates isolates which yielded nifH sequences on nifH PCR.

Phylogeny of potential diazotrophs

The phylogenetic tree of the isolates (Figure 3.6) shows the potential diazotrophs distributed according to various phyla of bacteria. The most abundant phylum was Proteobacteria, which further branch out into Alphaproteobacteria, Betaproteobacteria, and Gammaproteobacteria cluster. *Rhizobium* and *Phyllobacterium* were the most abundant Alphaproteobacteria. Isolates from the family *Burkholderiales* dominated the class Betaproteobacteria. *Variovorax* was the most abundant in Betaproteobacteria. *Pseudomonas* and *Stenotrophomonas* were the most abundant Gammaproteobacteria. Another gram-negative phylum isolated in this study was Bacteroidetes which was dominated by *Flavobacterium.*

Both gram-positive phyla, Actinobacteria and Firmicutes were detected. Actinobacteria was the most dominant with four clusters represented by *Streptomyces, Rhodococcus, Kribbella,* and *Microbacterium.* Firmicutes were mainly represented by *Bacillus* and very few *Paenibacillus* and *Staphylococcus.*

3.3.4 Survey of *nifH* gene among the isolates

Genomic DNA from all the isolates was tested for the presence of *nifH* gene using *nifH* primers listed in Table 3.1. Among the three primer sets, only polF/polR pair gave amplicons with our isolates. Only 81 isolates among 474 isolates yielded *nifH* sequences. The majority of *nifH* sequences were obtained from isolates identified as *Streptomyces* (30), *Microbacterium* (12), *Rhizobium* (5), *Rhodococcus* (4), *Stenotrophomonas* (4), *Paenibacillus* (4), *Nocardioides* (4), *Mesorhizobium* (2), and *Phyllobacterium* (2). The phylogenetic identification of these *nifH* sequences was performed by aligning them to

the reference *nifH* tree obtained in chapter 2. The *nifH* sequences clustered at the following four locations in the reference tree: (Figure 3.7)

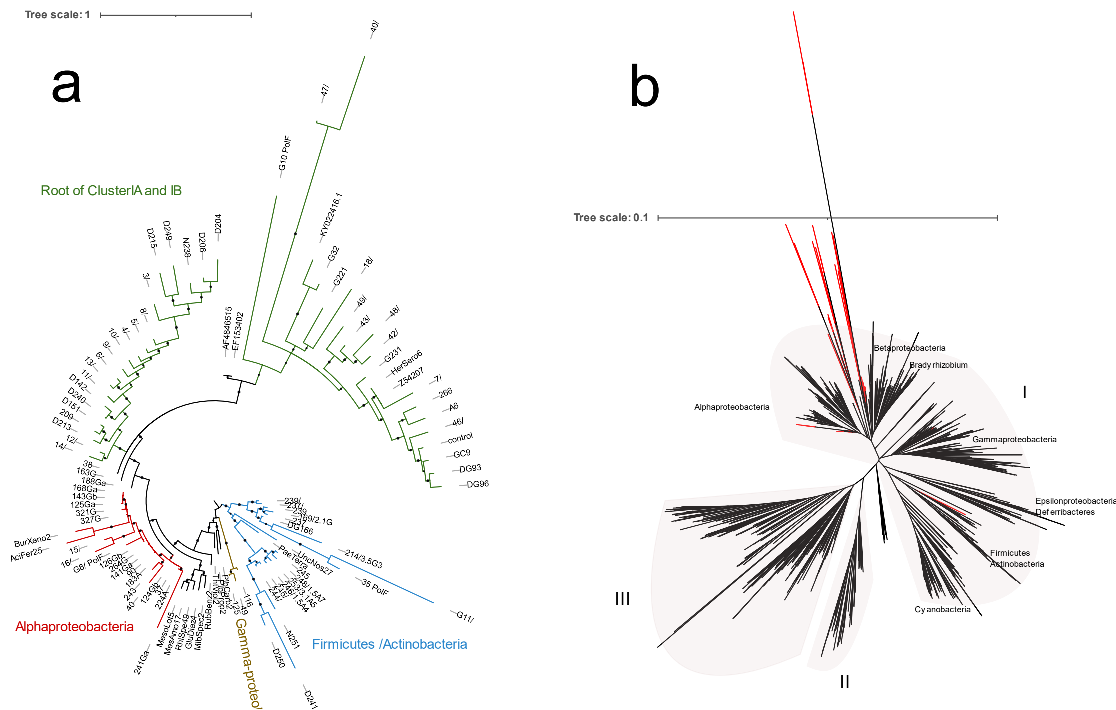


Figure 3. 7 (a)Phylogenetic analysis of the nifH sequences obtained in this study constructed using PhyML. Size of black dots on the brach represent the aBayes branch support. (b) Phylogeny using 926 nifH sequences obtained from previous study (chapter 2) as references. Tree constructed using RapidNJ with Kimura evolution model. The branches colored in red indicate sequences isolated in this study and nifH clusters are named according to (Zehr et al. 2003).

a. Alpha group: This cluster contains reference *nifH* genes from all the Alphaproteobacteria like *Rhizobium,* and *Mesorhizobium*, except *Bradyrhizobium.* 16 *nifH* from this study clustered here of which only one was identified as *Mesorhizobium* by 16S rRNA phylogeny. Three of the rest were identified as *Streptomyces* and the remaining 14 were identified as *Microbacterium* by 16S rRNA phylogeny.

b. Root of Alpha and Beta group: The majority of the *nifH* (44) obtained in this study clustered at the root of the Alpha and Beta group with very low

sequence similarity with the reference *nifH*, indicated by long branch lengths in the phylogenetic tree (Figure 3.7). 22 *nifH* from *Streptomyces,* 5 from *Rhizobium,* and other from genera like *Mesorhizobium, Pseudomonas, Phyllobacterium,* and *Stenotrophomonas* occurred in this cluster. None of these clustering is in harmony with the reference *nifH* sequences.

c. Gamma group: Three *nifH* obtained from *Stenotrophomonas* in this study occur in Gamma cluster which is in harmony with reference *nifH* sequences in this group.

d. Firmicutes group: This group contains reference *nifH* sequences from aerobic Firmicutes like *Paenibacillus* and Actinobacteria like *Frankia.* 13 *nifH* from this study aligned within this cluster. 11 *nifH* grouped here are from *Paenibacillus* (4), *Rhodococcus* (4), *Mycobacterium* (1), *Nocardioides* (1), and *Micromonospora* (1) which are in sync with the reference *nifH* in this group. While remaining two, one each from *Phyllobacterium* and *Mitsuaria* disagree with reference *nifH* in this group.

3.3.5 Quantification of nitrogen fixation by $^{15}N_2$ assimilation

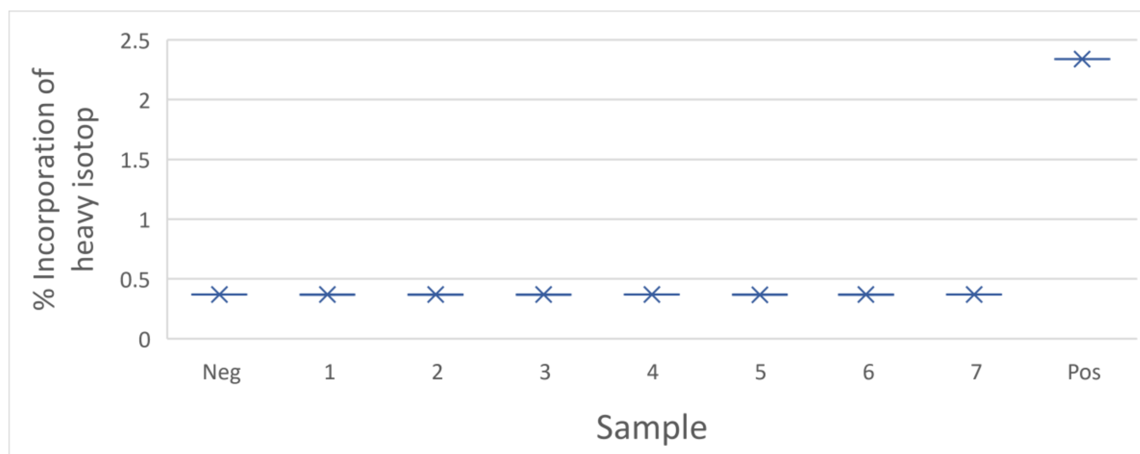


Figure 3. 8 $^{15}N_2$ assimilation of seven samples shows assimilation comparable to the negative control (E. coli). H. seropedicae was used as the positive control.

Of the 474 isolates, seven isolates from the phylum Actinobacteria were selected for $^{15}N_2$ assimilation based on the observance of growth rate in liquid NFM. A significant amount of incorporation of ^{15}N was observed only in the control, and samples had the same level of incorporation same as the negative control (Figure 3.8).

3.3.6 Quantification of nitrogen fixation by ARA

ARA was used to both confirm the diazotrophy and measure the rate of N_2 fixation for the positive isolates. Multiple isolates from the phylum Actinobacteria and Firmicutes were tested for ARA, however, only isolates of *Paenibacillus* were positive, and all the *Streptomyces* and *Microbacterium* were either negative or yielded inconsistent results (Figure 3.9).

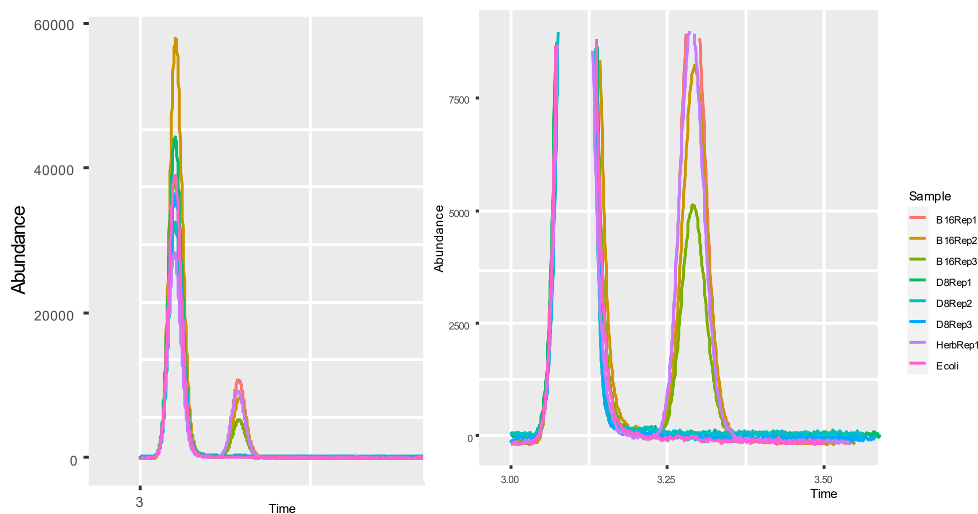


Figure 3. 9 Acetylene reduction assay of the selected isolates (left) and zoomed in (right) to show the ethylene peak. Presence of ethylene peak confirmed the diazotrophy.B16 (Paenibacillus), and D8 (Bacillus)are samples and Herb and E coli are positive and negative controls respectively.

3.4 Discussion

Significant diversity of free-living putative diazotrophic bacteria was observed in the soil samples obtained from Sioux prairie using a culture-based approach. The diversity-increasing approach described by (Mirza and Rodrigues 2012), where samples were inoculated on solid NFA plates instead of semisolid or nitrogen free broth, was used for isolation of potential diazotrophs. This method yielded on average, 10^8 CFU/gm of

potential free living diazotrophs which is more than 10^4 times the estimates by (Rilling et al. 2018), who estimated only 10^3 CFU/gm of putative N_2 fixing bacteria in the wheat rhizosphere using semisolid NFB. Similarly, other methods using semisolid media, irrespective of the sample used, reported low counts of putative nitrogen fixing bacteria. 10^4 CFU/gm and 10^4-10^5 CFU/gm of putative diazotrophs were reported in chickpea rhizosphere and sugarcane rhizosphere by (Pathani et al. 2014) and (Ahmad et al. 2006) respectively. The vastly higher isolation number is probably due to the spatial separation of slow growing stains from other fast-growing strains on the surface of agar, thereby allowing enough time for slow growers to form visible colonies on agar plate without being outcompeted by faster growers (Mirza and Rodrigues 2012).

On phylogenetic classification of partial 16S rRNA sequences, members of the phylum Proteobacteria (43%), Actinobacteria (42%), Firmicutes (6%), and Bacteroidetes (2%) were found in this study. (Rilling et al. 2018) also reported these four phyla from wheat rhizosphere using culture-based approach. Similarly, (Mirza and Rodrigues 2012) also reported isolates from these four phyla from Amazonian forest soil using a culture based approach. (Mirza and Rodrigues 2012) also made the comparison of bacterial diversity obtained using three approaches based on the effect of oxygen level and state of isolation media (solid/ semisolid). They found significant increase in diversity when plated directly on solid media and incubated under reduced oxygen condition, but we did not observe a significant difference between the CFU count under aerobic and reduced oxygen conditions. Also, most of the bacteria reported by Mirza and Rodrigues, 2012 only in reduced oxygen incubation were also present in aerobic incubation condition in

our case. Therefore, the inoculation of samples on an agar surface appears to be the primary reason for increased diversity rather than reduced oxygen condition.

Culture independent study of nitrogen-fixing bacteria associated with Switchgrass (Bahulikar et al. 2014) using *nifH* 3 *and nifH* 4 primers reported diazotrophic species from Alpha-, Beta-, Delta-, Gammaproteobacteria and Firmicutes only. Similarly, study of Tibetan grassland soils (Che et al. 2018) using PolF/R primers reported the species Cyanobacteria, Proteobacteria and Verrucomicrobium. The absence of diazotrophic Actinobacteria which represents nearly half of the isolates in culture-based approach is noteworthy and this could be the result of *nif* primer bias or because of inability of selective media to distinguish between true diazotrophs and oligotrophic bacterial species. There have been several reports of false positive diazotrophic fungi and bacteria because of their ability to grow in nitrogen free medium (Rosenberg et al. 2013). To address this issue, we eliminated all the components like yeast extract and vitamins from the selective media that could be a potential source of nitrogen and did the chemical analysis of the media for total nitrogen, ammonia, nitrite, and nitrate content. The nitrogen free medium in this study contained only 6.42, 1.19, 0.31, and 0.97 mmol/L of total nitrogen, ammonia, nitrite, and nitrate respectively. Nitrogen was not detected in the noble agar used for preparing the solid NFA. Also, to make sure no ammonia in the atmosphere supports the growth of bacteria, we incubated the isolates in sealed containers with Clinoptilolite, an ammonia scavenger (Tosun 2012). These strict measures should only encourage true diazotrophs to grow on NFA but the possibility of very efficient N-scavengers growing on our NFA cannot be eliminated completely.

Most of the isolates obtained in this study were well known diazotrophs (Table 3.2) except the isolates from Actinobacteria and Firmicutes. The most abundant bacteria in this study were *Streptomyces. S. thermoautotrophicus* was reported to fix nitrogen by an oxygen insensitive novel nitrogenase, which did not reduce acetylene to ethylene (Gadkari et al. 1992). This claim was refuted by another group of researchers who failed to observe nitrogen fixation in *S. thermoautotrophicus* (MacKellar et al. 2016). However, other *Streptomyces* are still frequently reported to fix nitrogen or possess plant beneficial roles (Le et al. 2016; Sellstedt and Richau 2013). Recently, an agar degrading, nitrogen fixing *S lavendulae* was reported to contain possible nitrogen fixation gene *nifU* (Wu et al. 2017). The second most abundant Actinobacteria in this study, *Microbacterium* has also been previously isolated in nitrogen free medium (Rilling et al. 2018) and reported to fix nitrogen (Gtari et al. 2012), but still lacks genomic evidence. Species of *Rhodococcus* are also frequently isolated in nitrogen free medium (Blasco et al. 2001; Joshi et al. 2019a) and their ability to scavenge atmospheric NH_3 has be studied in detail (Yoshida, Inaba, and Takagi 2014). Similarly, *Mycobacterium* has been reported to fix nitrogen (Gtari et al. 2012) but a novel glucosylglycerate hydrolase enzyme has been suggested responsible for the recovery of *M hassiacum* from nitrogen starvation (Alarico et al. 2014). Similarly, there have been multiple anecdotal evidence of diazotrophy in other species of Actinobacteria like *Agromyces, Williamsia,* and *Kocuria* but the phylogenetic analysis of all the diazotrophs (Chapter 2), suggest that *Frankia* and *Propionibacterium* are the only two Actinobacteria with a complete minimal set of nitrogenase genes in the sequenced genome, and supported by biochemical evidence.

Bacillus was the most abundant Firmicutes in this study, but the taxonomy of the genus *Bacillus* has undergone multiple revisions and diazotrophy of *Bacillus* needs some clarification. Most *Bacillus* found to have *nif* genes, such as *B. polymyxa* and *B. macerans*, have been reclassified as *Paenibacillus,* and diazotrophy among aerobic endospore-forming Firmicutes is restricted to the genus *Paenibacillus* only (Achouak et al. 1999). *B*. *nealsonii* and B *caseinilyticus*, the only *Bacillus* that possess all core six *nif* genes, are extremophiles and lack biochemical evidence for nitrogen fixation (chapter 2). This is in contrast to multiple claims of diazotrophic *Bacillus* isolates, supported only by culture evidence for growth on nitrogen free media (Ding et al. 2005; Guo et al. 2020; RoÂzÇycki et al. 1999; Yousuf et al. 2017). However, none of these claims have been supported by genetic evidence. Most of the other isolates belonging to Proteobacteria isolated in this study are well known diazotrophs or phylogenetically closely related to diazotrophs.

The recovery rate of *nifH* sequences from the isolates using PolF/R primer set was only 17% which is very low compared to 79% *nifH* positive isolates reported by (Mirza and Rodrigues 2012). The inability to recover *nifH* sequence from pure cultures has been observed in other studies as well (Achouak et al. 1999; Berg et al. 2002; Doty et al. 2009). This suboptimal performance of *nifH* PCR is partially due to the less conserved nature of the *nifH* gene. The *nifH* gene varies considerably across the known diazotrophs and does not have conserved region consensus among all the diazotrophs. This has led to the design of multiple *nifH* PCR primer sets optimized to specific groups of bacteria (Zehr and Capone 2020). Also all the primers designed so far are all degenerate, decreasing the specificity of the amplification (Gaby and Buckley 2012a). Most of the

nifH sequences obtained in this study were not homologs to the 16S phylogeny. All of the *nifH* sequences obtained from *Streptomyces* and *Microbacterium* aligned with the *nifH* from Proteobacteria. Horizontal gene transfer (HGT) is common among nitrogenase genes (Raymond et al. 2004), but no genomic evidence of HGT has been observed between Actinobacteria and Proteobacteria (Chapter 2). Therefore, the most likely explanation for this, could be a co-culture of *Streptomyces* with Alphaproteobacteria. Several *Streptomyces* and Actinobacteria have been reported to be isolated from root nodules and shown to increase nitrogen fixation when *Streptomyces* and rhizobia are co-inoculated to plants (Le et al. 2016; Rilling et al. 2018).

3.5 Conclusion

The diversity-increasing approach to culture potential diazotrophs, using solid NFA, solidified with noble agar increases the diversity of culturable diazotrophic bacteria. Conversely this approach yielded a high proportion of isolates without *nifH.* Representatives from Proteobacteria, Actinobacteria, Firmicutes and Bacteroidetes could be isolated and majority of them have both biochemical and genetic evidence to support diazotrophy. However, genus-wise distribution shows dominance of a single genus, *Streptomyces* which needs further study as these bacteria are frequently reported to be isolated in NFM, but no concrete biochemical and genetic evidence is available to support diazotrophy in this genus to date.

Chapter 4

Characterization of *Streptomyces* Isolate Growing in Nitrogen Free Medium and Study of Differentially Expressed Genes in Nitrogen Deplete and Replete Condition.

4.1 Introduction

Nitrogenase is the only enzyme system known to occur in a diverse taxon of prokaryotes called diazotrophs which can reduce the dinitrogen available abundantly in the atmosphere to the reduced form (NH_3) that can be assimilated by plants and other microbes (Postgate et al. 1982). Nitrogenase is a complex enzyme made up to two components (dinitrogenase (heterotetrmer of NifD and NifK) and dinitrogenase reductase (homodimer of NifH) (Hu and Ribbe 2015). The most commonly used marker for studying diazotrophy is *nifH,* which has been used to identify this diverse range of prokaryotes with the potential to fix nitrogen (Zehr et al. 2003). Based on the lifestyle adopted, diazotrophs are classified as symbiotic (forming root nodules with higher plants like *Rhizobium, Bradyrhizobium, Mesorhizobium, Nostoc,* and *Frankia*) and free-living (fixing nitrogen independent of higher plants like *Azotobacter, Pseudomonas, Klebsiella, Azospirillium, Paenibacillus,* etc.). Although free living diazotrophs are less efficient in fixing dinitrogen compared to symbiotic diazotrophs, their ubiquitous distribution is thought to play a vital role in the global nitrogen cycle (Vitousek et al. 2013).

Soil bacteria from the phylum Actinomyces are one of the most abundant bacteria found in the rhizosphere and the bulk of surface soil (Ventura et al. 2007). They are

exposed to a diverse range of nutrients and growth conditions which resulted in evolution of a vast range of metabolic capabilities and functions in these bacteria (Krysenko et al. 2021). *Streptomyces* is a filamentous, Gram-positive, non-motile, spore forming, aerobic bacterium with guanine+cytosine (GC) content of 63-78% (Bentley et al. 2002). They are known to produce several plant growth promoting traits like support nitrogen fixation and solubilization of minerals (zinc, potassium and phosphorus), chelation of iron, plant hormones and pathogen inhibiting capabilities (Gopalakrishnan, Srinivas, and Prasanna 2020). They produce a range of bioactive molecules like cellulase, lipase, protease, lipase, chitinase, indole acetic acid and antibiotics (Borodina, Krabben, and Nielsen 2005; Challis and Hopwood 2003). *Streptomyces* account for 39% of the medically applied secondary metabolites produced by Actinobacteria which include anti-fungal, anti-bacterial, antiviral, anticancer agents, insecticides, herbicides, antiparasitic agents, immune suppressants, antioxidants , and enzyme inhibitors (Bérdy 2012).

Several representative of the phylum Actinobacteria (*Frankia, Propionibacterium*) are well documented to fix atmospheric nitrogen and occur in symbiotic association with plants (Gtari et al. 2012; Trujillo et al. 2015). Also, some species of Actinobacteria has been isolated from root nodules of nitrogen fixing plants and shown to possess plant beneficial effects (Le et al. 2016; Rilling et al. 2018). In addition to this, several free living Actinobacteria are reported to fix nitrogen (Sellstedt and Richau 2013). *Streptomyces* is one of the most commonly isolated bacteria in nitrogen free medium (Mirza and Rodrigues 2012; Rilling et al. 2018). A unique, oxygen insensitive nitrogenase has been reported in *S. thermoautotrophicus* which is not inhibited by dinitrogen analogues like acetylene, ethylene, CO and H_2 (Gadkari et al. 1992). However,

the existence of such nitrogenase is in question because another group has refuted diazotrophy in *S. thermoautotrophicus* (MacKellar et al. 2016). An agarivorans actinobacteria, *S. lavendulae* was reported to have potential nitrogen fixing ability because of presence of *nifU* gene (Wu et al. 2017). However, the thorough examination of 561 complete genomes of Actinobacteria available in NCBI indicates only *Franki* and *Propionibacterium* have the complete set of nitrogenase enzyme (Chapter 2). Hence, diazotrophy in the genus *Streptomyces* is not clear and needs rigorous biochemical and genomic screening.

Streptomyces sp. 321G was isolated in our lab on a very stringent nitrogen free agar (NFA) (Chapter 3). Its growth in nitrogen free medium was confirmed in both NFA (by re-streaking) and in liquid nitrogen free medium by incubation in an ammonia free closed chamber. *nifH* PCR on the genomic DNA of this isolate yielded a *nifH* sequence which was homologous to the *nifH* from *Herbaspirillium seropedicae* and the preliminary test with Acetylene Reduction Assay was also positive. These observations strongly suggest the nitrogen fixing ability of this isolate but there is no report of diazotrophic *Streptomyces* with concrete biochemical and genetic evidence. Hence this chapter focuses on characterizing the diazotrophic potential of this isolate.

4.2 Materials and methods

4.2.1 Bacterial specimen

Streptomyces sp. 321G was isolated from the bulk soil from Sioux prairie (44°01'55.2"N 96°46'59.6"W) located north of Colman in South Dakota. Of the 104

Streptomyces isolated in that study, 321G had the highest growth rate in nitrogen free medium and was one of the few isolates that was positive for *nifH* PCR.

4.2.2 Growth in nitrogen free medium

The growth potential of 321G was confirmed again by re-streaking on Nitrogen Free Agar (NFA) and growing them in liquid nitrogen free medium while incubating in ammonia depleted atmosphere by adding Clinoptilolite to the airtight incubation chamber (Tosun 2012). NFA media comprises of sugars (Glucose (2g/l), Arabinose (2g/l), Mannitol(2g/l), Malic acid (2g/l)) and K_2HPO_4 (0.2g/l), KH_2PO_4 (0.5g/l), $MgSO_4.7H_2O$ (0.2g/l), $FeSO_4.7H_2O$ (0.1g/l), $Na_2MoO_4.2H_2O$ (0.005g/l), NaCl (0.2g/l) and solidified with noble agar (15 g/l, Difco, Catlog No. 214230). To ensure the uniform inoculum size and prevent transfer of residual nitrogen through inoculum, 321G spores, washed three times with NFM were used for inoculation of all further experiments. Spores were harvested from three weeks old 321G cultures on R2A plates using cotton filters as described in https://openwetware.org/wiki/Streptomyces:Protocols/Spore_Prep. *E. coli K*12 and *Herbaspirillium seropedicae* ATCC 35892 were used as negative and positive controls respectively for all growth comparisons.

4.2.3 Microscopy

Samples for microscopy were obtained from liquid nitrogen free medium. Culture flasks were gently vortexed to obtain homogenous suspensions of mycelium, and 20 µl of culture was pipetted on a clean, grease free glass slide which was simple stained with 1% crystal violet and observed under bright field using Olympus BX-53 upright compound microscope.

4.2.4 Growth rate measurement

The cultures in NFM were quantified by measuring the amount of crystal violet binding. This method was adapted with some modifications from a rapid method developed by Fischer and Sawers for measuring growth of filamentous microorganisms (Fischer and Sawers 2013). Instead of methylene blue, 0.1% crystal violet was used for staining the mycelium pellets.

4.2.5 Extraction and sequencing of complete genome

Genomic DNA was extracted using CTAB (Ausubel 1988) and sequenced in the laboratory using MinION from Oxford Nanopore Technologies according to manufacturer's guideline. The raw fastq reads were assembled in galaxy (www.usegalaxy.org) using Flye assembly (Lin et al. 2016). The assembled genome was annotated using Prokaryotic genome annotation (Prokka) (Seemann 2014). The transcripts were also functionally annotated using KEGG BlastKOALA (Kanehisa, Sato, and Morishima 2016). To map the KO ids generated to microbial pathways, KEGG mapper was used (Kanehisa and Sato 2020).

4.2.6 Extraction and sequencing of total RNA

Total cellular RNA was extracted from strain 321G growing under three different conditions (NFM+NH_3, NFM (8 h), NFM (24 hours)), preparing three biological replicates for each condition. *Streptomyces* spores were inoculated into NFM+NH_3 medium and exponentially growing vegetative cells, washed with NFM for three times were used as inoculant for the three respective media and incubated at 28°C for the time according to the given condition. Five ml of culture were harvested to which 10 ml of

RNA protect Bacteria Reagent (Qiagen) was added for stabilization of RNA. Total RNA was extracted using the RNeasy Mini Kit (Qiagen) according to the product's manual. The RNA extract was quantified using a Qubit 3 Fluorometer and NanoDrop^{TM}2000/2000C Spectrophotometer. The extracts which passed the minimum requirement of 2.5 μg, were processed for sequencing to Novogene Corporation for Illumina paired end sequencing.

4.2.7 Processing and analysis of sequencing data

The reads were aligned to the *Streptomyces* 321G genome using TopHat2 (Kim et al. 2013) and aligned reads were counted with featureCounts (Liao, Smyth, and Shi 2014) to generate the count file. DESeq2 package in R was used for the identification of differentially expressed genes (Love, Huber, and Anders 2014).

4.3 Results

4.3.1 Culture characteristics

The preliminary evidence of potential to fix nitrogen by the isolate 321G was its ability to grow in nitrogen free medium: NFA and liquid nitrogen free medium.

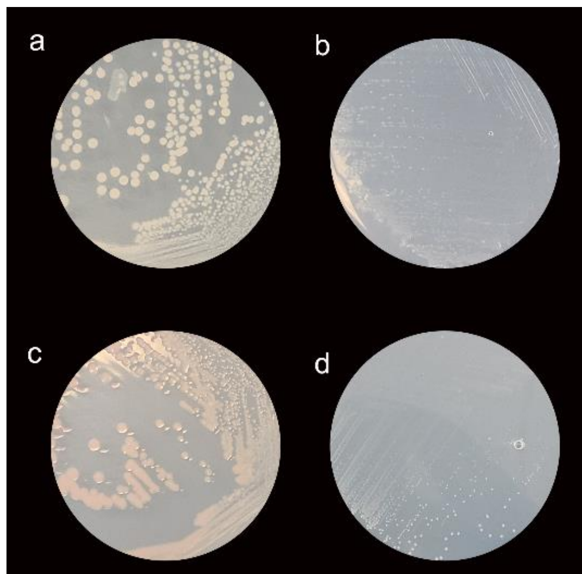


Figure 4. 1 Colony characteristics of Herbaspirillium seropedicae and Streptomyces sp. strain 321G on NFA for 3 d and NFA supplemented with NH_3 for 24 hrs . Strain 321G in NFA+NH_3 (a) and NFA (b) and H. seropedicae in NFA+NH_3 (c) and NFA (d).

4.3.1.1 Colony characteristics on agar

Colonies of 321G could be observed on NFA after 3 d of incubation at 28°C under aerobic conditions, where they produced small (1 mm in diameter), opaque, white, and leathery colonies with irregular margin and were partially embedded in the agar. On subsequent incubation for 14 d, the colonies increased in size, with wrinkled surface and granular deposits indicating sporulation. On NFA supplemented with ammonia colonies of size 1-2 mm were visible after 24 h of incubation at 28°C. Colonies on NH_3 containing media were also white and leathery with smooth margin and formed a wrinkled surface on longer incubation. The colony sizes and growth time in NFA were like the positive

control *H. seropedicae* (Figure 4.1). The negative control *E. coli* did not produce any visible colonies on NFA after 3 d of incubation.

4.3.1.2 Colony characteristics on liquid nitrogen free media

The ability of strain 321G to grow in nitrogen free medium was further confirmed by its growth in liquid nitrogen free medium incubated in an ammonia depleted closed chamber. Both strain 321G and *H. seropedicae* formed similar colonies after 14 d of incubation. The growth in liquid medium was characterize by the production of characteristic purple colored pigment for strain 321G. The cultures in NFM did not produce uniform turbidity and sedimented at the bottom of the flask for both strain 321G and the positive control (Figure 4.2).

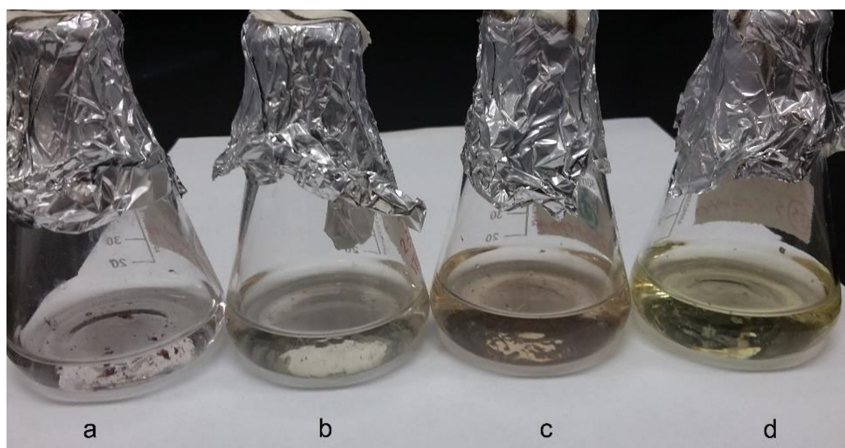


Figure 4. 2 Culture characteristics of strain 321G in NFM+NH_3 (a) and NFM (b) and H. seropedicae in NFM+NH_3 (c) and NFM (d).

4.3.2 Microscopy

Strain 321G produced extensive vegetative mycelium of branched hyphae, with a prominent septum separating the adjacent cells in NFM+NH_3 medium after 24 h of growth (Figure 4.3). Cells growing in excess nitrogen were ~5 μm in length. In nitrogen free liquid medium, there was no visible change in morphology after 8 h except for slight

shrinkage of cells indicated by the unstained region on the either side of the septum. After 24 h of incubation on NFM, the hyphae appeared to be fragmented and there was decrease in cell size as well. The cell shape also changed from elongated rods to oval after 24 h of incubation. After 7 d of incubation on NFM, there was significant change in cell morphology. The hyphae appeared to be a chain of oval cells indicating the formation of endospores.

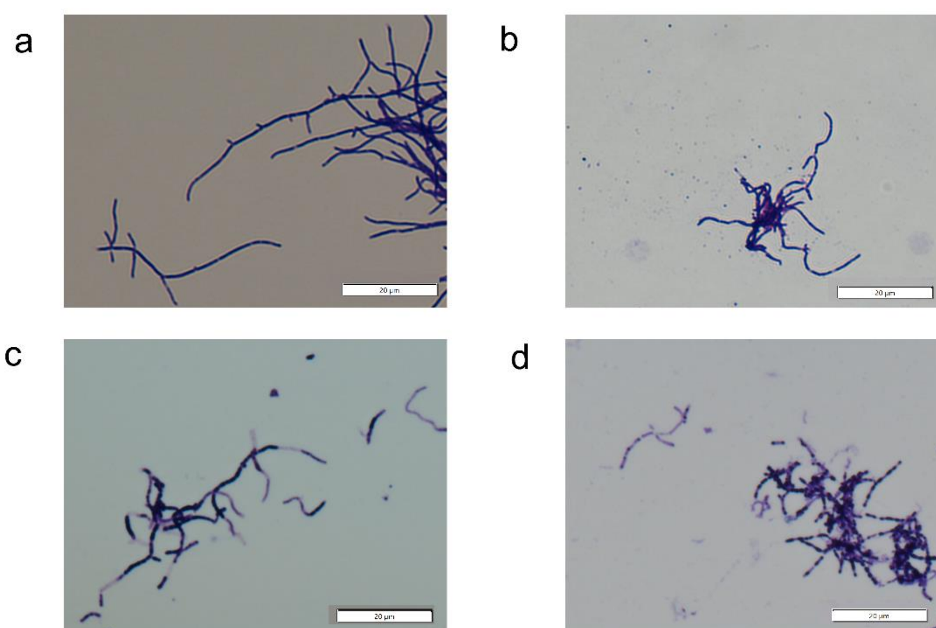


Figure 4. 3 Streptomyces filaments under bright field microscope after staining with crystal violet. Streptomyces spores growing in NFM+NH_3 media after 24 h (a). Exponentially growing filaments in NFM+NH_3 transferred to NFM for 8 h (b), 24 h (c) and 7 d (d).

4.3.3 Growth characteristics

Strain 321G showed increase in biomass in both nitrogen replete and deplete condition however the growth rate was much slower in absence of nitrogen (Figure 4.4). In presence of ammonia, the doubling time was 24 h and cells entered stationary phase

after 4 d. In absence of nitrogen, doubling time was more than 48 h. The biomass produced at the end of 5th day in nitrogen free medium was 12 times less than that in nitrogen rich medium. However, there is also a 12-fold increase in biomass in nitrogen free medium and cells did not enter the stationary phase when growing in nitrogen free medium for 5 d which indicates the slow but consistent growth.

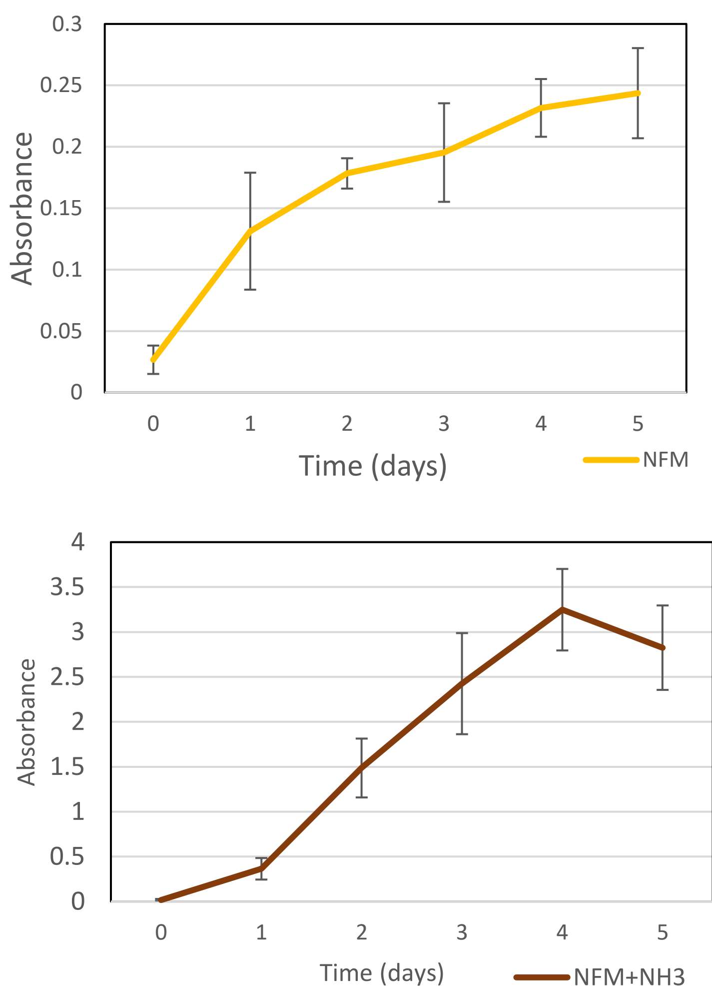


Figure 4. 4 Growth of Streptomyces sp. strain 321G estimated by absorbance by crystal violet after elution of the crystal violet bound to cell pellets in NFM (Top) and NFM+NH_3(bottom).

4.3.4 Measurement of nitrogen fixation

Strain 321G was tested for its nitrogen fixation ability by acetylene reduction assay (ARA), $^{15}N_2$ incorporation assay, and *nifH* PCR. ARA was inconclusive as an ethylene peak was detected only once out of three runs (data not shown). $^{15}N_2$ incorporation assay showed no significant enrichment of the biomass with heavy isotope of nitrogen. However, *nifH* PCR using polF/R primers was positive and yielded a *nifH* sequence homologous to the *nifH* from *H. seropedicae.*

4.3.5 Analysis of *Streptomyces* genome

The complete genome of strain 321G was sequenced using the MinION Oxford Nanopore technology, generating 106 fastq files containing 4000 reads in total. The average GC ratio was 68%. The reads were checked for a possible biculture by analyzing the taxonomy of each read. Of the total number of reads, 79% of reads were found to be from the genus *Streptomyces* and 12% were classified as other Actinomyces. Three percentage of reads belonged to a Gammaproteobacteria, *Lysobacter* and the remaining 6% could not be classified as any prokaryotes.

The sequence data generated by MinION was assembled using Flye assembly into 18 contigs of varying length (3,561 to 11,593,334 bp) with total length of 17.5 Mbp. After the annotation of the assembled reads with Prokka, eight 16S rRNA sequences were detected of which, 6 aligned with *Streptomyces phaeochromogenes* and two aligned with the Gammaproteobacteria, *Lysobacter sp* (Figure 4.5).

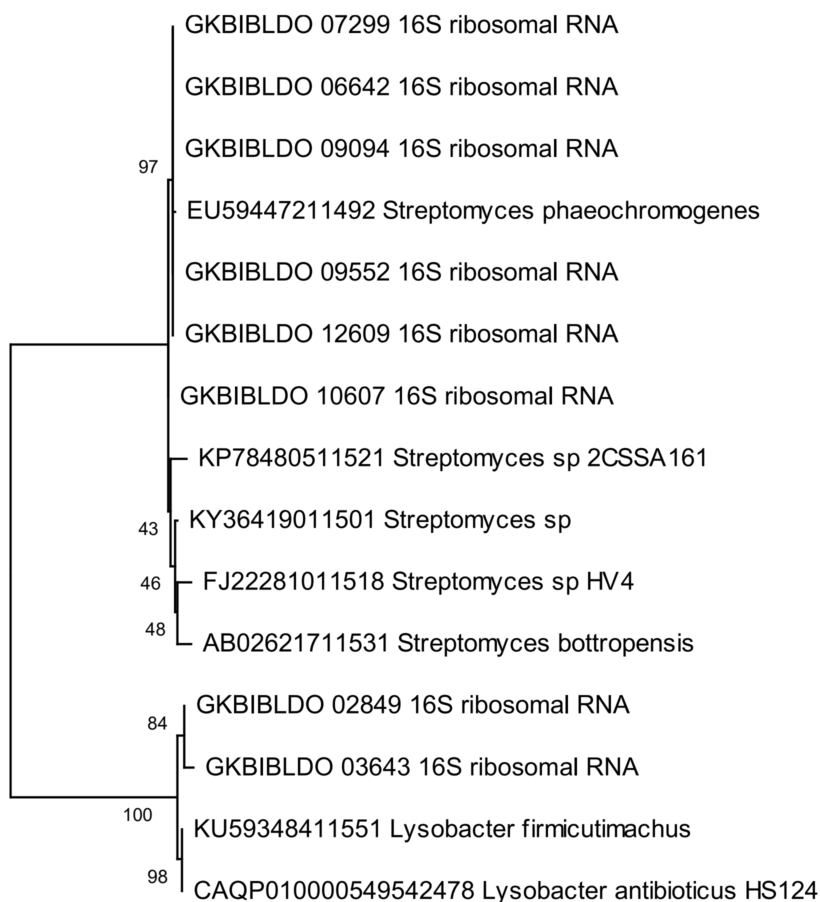


Figure 4. 5 Molecular phylogenetic tree of eight 16S rRNA sequences obtained from the assembled reads determined by Maximum-likelihood with 100 bootstrap replicates. The reference sequences are selected based on the closest alignment in SINA database.

Prokka predicted 18,759 transcripts in total of which only 3,718 could be annotated with KO terms using KEGG KOLA annotation pipeline. It identified genes involved in metabolism, genetic information processing, environmental information processing and some unclassified genes not included in Brite pathway (Figure 4.6). Majority of genes involved in metabolism were for carbohydrate, amino acid, vitamins and co-factors metabolism, and xenobiotics biodegradation and metabolism. Genes involved in replication and repair and translation were most abundant in Genetic information processing category. Membrane transport were the most abundant in the Environment Information Procession category.

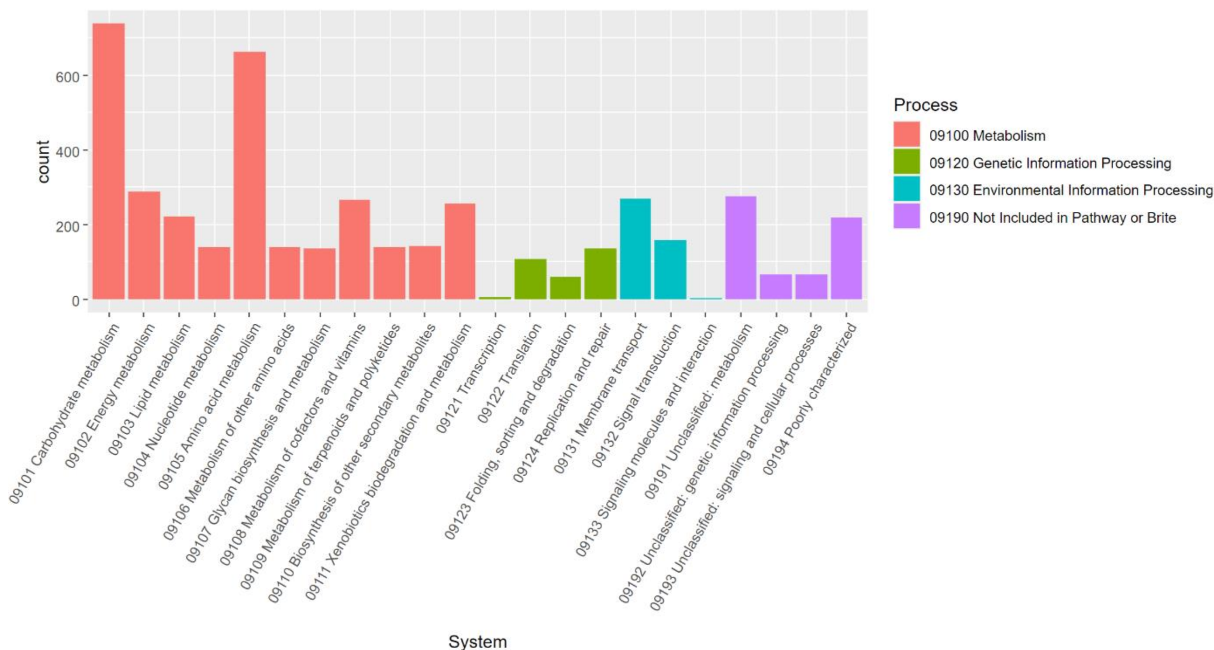


Figure 4. 6 Functional classification of genes identified in the strain 321G. The genes were annotated and mapped by KEGG.

In addition to central carbohydrate metabolism (glycolysis, gluconeogenesis, pyruvate oxidation, citrate cycle, pentose phosphate pathway), this strain also possesses complete pathways for other carbohydrate metabolism like galactose degradation, nucleotide sugar biosynthesis, and the glyoxylate cycle. For carbon fixation, complete

pathways were present for reductive pentose phosphate cycle, CAM (Crassulacean acid metabolism) and phosphate acetyltransferase-acetate kinase pathway. No gene involved in nitrogen fixation (*nifHDKENB)* was detected by KEGG (confirmed by manual NCBI Blast search as well), however a partial pathway for nitrate assimilation was detected by KEGG. A complete pathway was detected for assimilatory sulfate reduction and ATP synthesis. A complete pathway was also detected for fatty acid metabolism (fatty acid biosynthesis, beta-oxidation), and lipid metabolism (Phosphatidylcholine (PC), Phosphatidylethanolamine (PE) biosynthesis). Multiple complete pathways were also present for purine and pyrimidine metabolism. This strain also had complete multiple pathways for metabolism of cofactors and vitamins (Pyridoxal-P biosynthesis, NAD biosynthesis, Coenzyme A biosynthesis, Biotin biosynthesis, Lipoic acid biosynthesis, Tetrahydrofolate biosynthesis, etc.).

4.3.6 Differential gene expression

4.3.6.1 RNA extraction and quantification

Total cellular RNA was extracted from strain 321G growing under three conditions: NFM supplemented with ammonia incubated for 24 h, NFM incubated eight h and NFM for 24 h. The first condition is the control against which differentially expressed genes from the other two conditions were compared (Table 4.1). All the RNA extracts were quantified by Fluorometric method (Qubit) and RNA for Illumina library preparation and sequencing was found to be sufficient.

Table 4. 1 RNA quantities from nine samples passing the necessary requirements for sequencing.

Sample	RNA concentration (μg/ml)	Description
C1	198	NFM+NH3
C2	260	NFM+NH3
C3	279	NFM+NH3
S1	216	NFM (8 hours)
S2	224	NFM (8 hours)
S3	159	NFM (8 hours)
S4	107	NFM (24 hours)
S5	147	NFM (24 hours)
S6	72.8	NFM (24 hours)

4.3.6.2 Read quality

On average 10 million paired end reads of length 150 bp were sequenced from each sample and the GC ratio was 69% for all the samples, which corresponds to the GC content of *Streptomyces.* On average there were 70% duplicate reads indicating good coverage of the total mRNA in the sample. Also, the phred quality score for all the samples was above 30 indicating good sequencing quality (Figure 4.7).

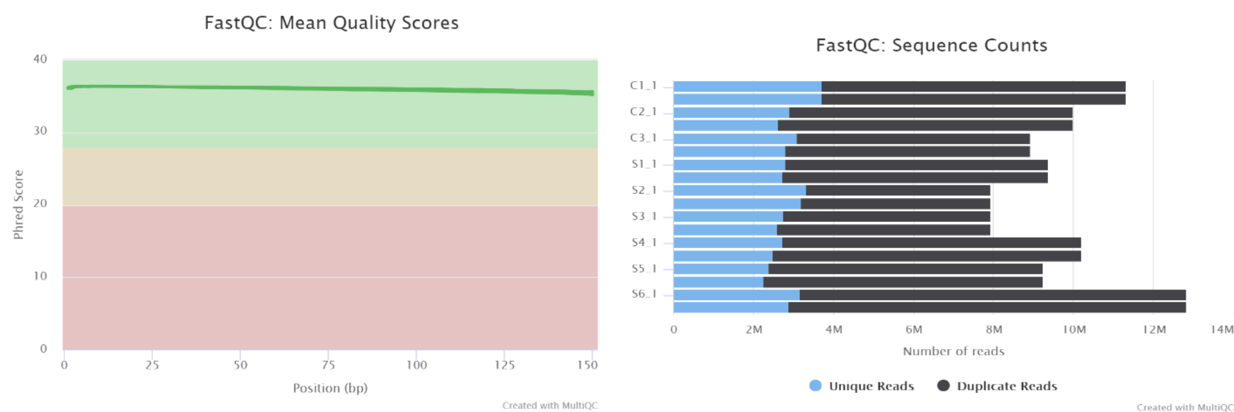


Figure 4. 7 The quality score and sequence count on each samples, obtained from fastQC on the raw fastq files.

4.3.6.3 Taxonomy of reads

The taxonomy of all the raw reads were also obtained by Kaiju. For all the samples, 88% of the reads were identified as *Streptomyces* except for C1 for which only 82.6% of reads were identified as *Streptomyces.* On average 9% of other reads were identified as some other Actinomyces and remaining reads were unclassified (Figure 4.8). This indicates most of the samples had very little contamination and sample C1 appears to be slightly different than other samples.

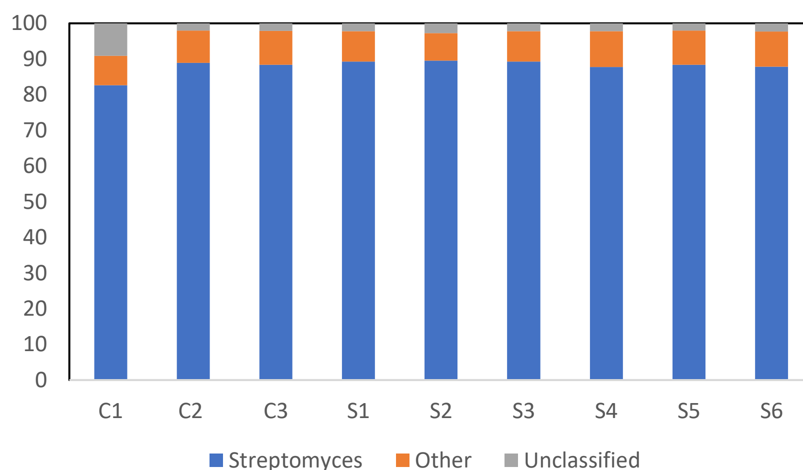


Figure 4. 8 Taxonomy of raw fastq reads obtained from the nine samples given by Kaiju. Sample names are defined in Table 4.1.

4.3.6.4 Sample clusters

All the raw reads were aligned against the strain 321G genome sequence obtained from nanopore sequencing. Of the 18,759 genes present in the reference genome only, 5,027 genes had at least one read aligned to it, indicating that the strain 321G has vast metabolic capabilities and only 26.79% genome is essentially necessary for growth in minimal media with and without nitrogen.

The distances between the samples are plotted according to the expression profile of the genes (Figure 4.9). Figure 4.9a shows sample C1 has very different expression

profile than any other samples, which might affect the differential gene analysis hence the expression from C1 was not used for the further analysis. Figure 4.8b shows sample distance without C1 and shows the clustering of samples according to the growth condition and PCA analysis (Figure 4.9c) shows clustering of samples according to the experimental conditions when C1 was not considered on the analysis.

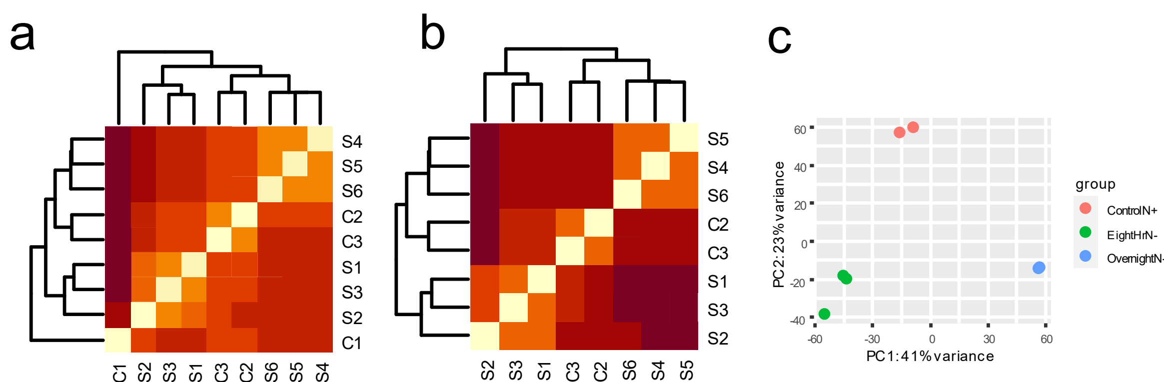


Figure 4. 9 Sample distance (a, b) and principal component analysis (c) of the expression profile according to the different sampling condition.

4.3.6.5 Differentially expressed genes

DESeq2 analysis showed many genes were significantly expressed differentially when comparing control with NFM (8 hr) or NFM (24 hr) (Figure 4.10). Out of the total 5,027 expressed genes 1,130 were expressed differentially between control and NFM (8 hr) and 1,034 were differently expressed between control and NFM (24 hr). Of these differentially expressed genes 485 were common between the two nitrogen deficient conditions and their expression patterns were similar in these two conditions.

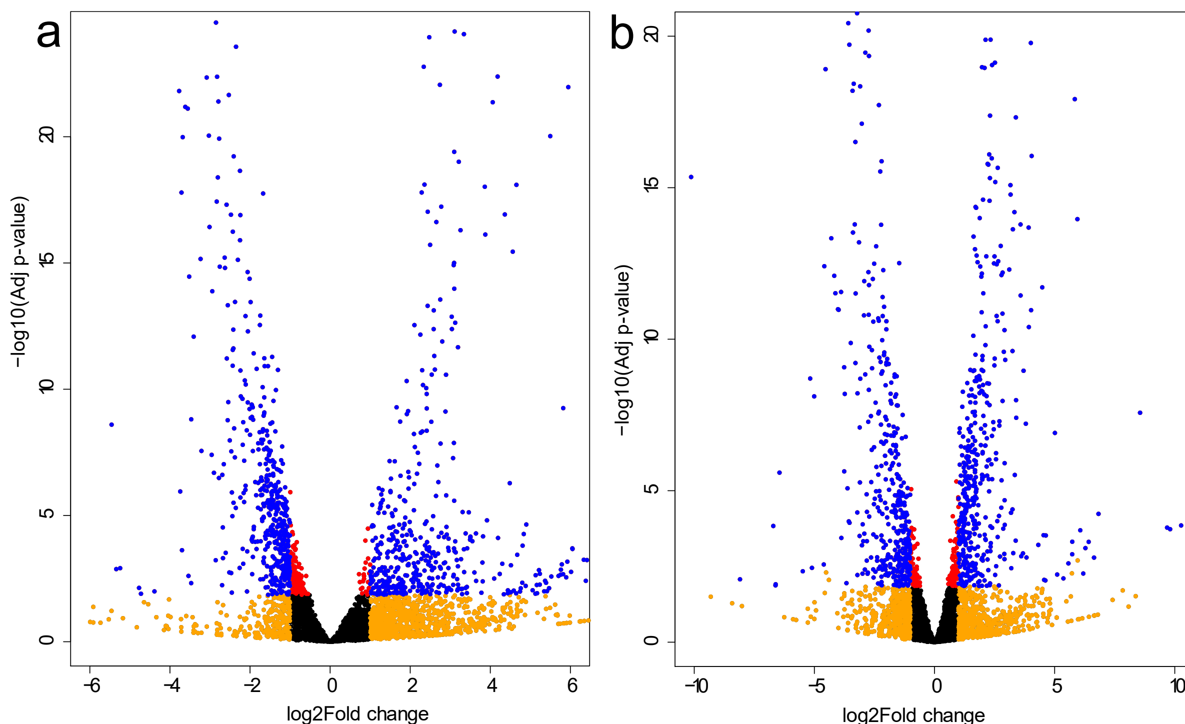


Figure 4. 10 Volcano plots showing differentially expressed genes by log_2fold change and -log10(Adjusted p-value) between control Vs NFM(8 h) (a) and control Vs NFM (24 h) (b). Genes having adjusted p-value less than 0.05 and log_2fold change greater than 1 are colored blue.

4.3.6.6 Functional analysis of the differentially expressed genes

The expression pattern of different functions in both 8 h and 24 h incubation in NFM appears to be very similar with few differences in some pathways (Figure 4.11). The exposure of these cells to NFM caused change in expression of multiple metabolic, genetic information processing, environmental information processing, and non-Brite pathways. However, the eight-h exposure was characterized by increase in expression of membrane transport and signal transduction genes and 24 h exposure was characterized by increase in amino acid metabolism, xenobiotics metabolism, transcription, and replication and repair of nucleic acid. The following pathways had more than 2 (upregulated) or less than -2 (downregulated) log_2fold change in each biological process:

Carbohydrate metabolism: Genes related to glycolysis/gluconeogenesis, pyruvate metabolism, citrate cycle, butanoate metabolism were downregulated in both conditions but amino sugar and nucleotide sugar metabolism were upregulated in both conditions. Starch and sucrose metabolisms were upregulated in 8hr condition.

Energy metabolism: Genes related to methane metabolism and carbon fixation were downregulated and Nitrogen metabolism was upregulated in both conditions. Glutamine synthetase, nitrate/nitrite transporters were the proteins upregulated in nitrogen metabolism in both conditions, which increased up to 10 log_2fold in the 24h condition and 7 log_2fold in 8 h condition.

Lipid metabolism: Multiple genes related to fatty acid degradation and synthesis of unsaturated fatty acids were both upregulated and downregulated in both conditions. Higher proportion of genes related to lipid metabolism were upregulated in 8 h condition, but higher proportion of lipid metabolism genes were downregulated in 24 h condition.

Amino acid metabolism: A higher number of genes related to amino acid metabolism were either upregulated or downregulated in the 24 h condition compared to the 8 h condition. Urease, glutamine synthetase, phenylacetyl-CoA epoxidase was upregulated, and acyl-CoA dehydrogenase were downregulated in both conditions.

Glycan metabolism: Beta-N-acetylhexosaminidase, beta-galactosidase and alpha-1,6-mannosyltransferase were upregulated in the 8 h condition and phosphomannomutase, penicillin-binding protein A were downregulated in the 24 h condition.

Genetic information processing: Transcription is increased in both conditions, but 24 h had more than 2 log2fold increase. Translation was increased in both conditions, shown by upregulation of aminoacyl-tRNA biosynthesis. The eight-hour condition was characterized by downregulation of the sulfur relay system related to folding, sorting and degradation of nucleic acids. The 24 h condition showed upregulation of genes related to DNA replication and repair.

Membrane transport and signal transduction: The 8 h condition was characterized upregulation of genes related to membrane transport. All the upregulated transport systems are ABC transporters and related to uptake of oligopeptides, ribose, branched-chain amino acids, and sugars. A two-component signal transduction system, nitrogen regulatory protein P-II and Bifunctional uridylyltransferase/uridylyl-removing enzyme were upregulated in both conditions.

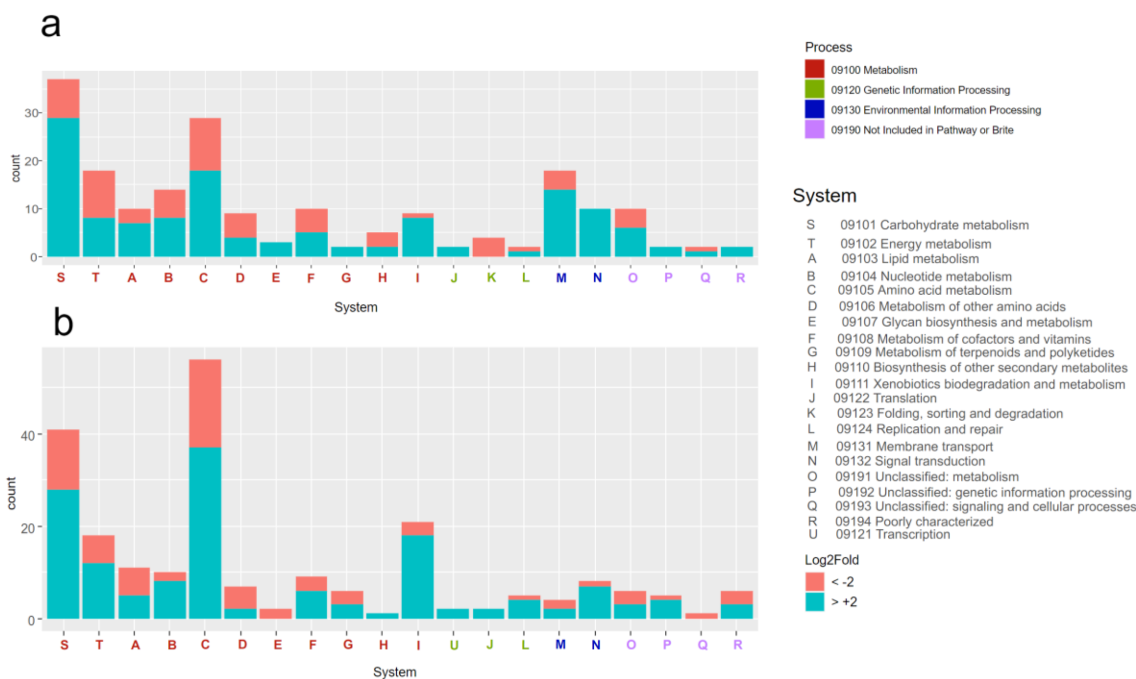


Figure 4. 11 Functional classification of differentially expressed genes in eight-hour condition (a) and 24-hour condition (b). Only the genes with more than 2 or less than -2 log$_2$fold change are presented.

Hypothetical proteins: Besides genes annotated by KHOLA, 47.5% of differentially expressed genes in eight-hour condition and 67.7% genes in 24-hour condition are hypothetical proteins without clear functional classification. Prokka also didn't annotate 41.11% of differentially expressed genes in eight-hour condition and 43.9% genes in 24-hour condition. Some of these hypothetical proteins had trans-membrane helix, ABC transporters, and other domains related to the amino acid and cofactor metabolism but any domain with potential role in nitrogen fixation could not be identified. Larger fraction of hypothetical proteins didn't have any previously identified domain.

4.4 Discussion

Streptomoyces strain 321G was isolated from the prairie soil on a nitrogen free agar and growth on nitrogen free medium was again confirmed by growing the isolate on NFA

and in liquid NFM in a closed chamber with depleted ammonia. The measurement of growth rate of this strain in NFM showed 12-fold increase in biomass. *Streptomyces* are one of the most abundant genera in rhizosphere soil and play vital role in growth and overall health of plant by producing various plant growth promoting (PGP), plant health promotion effects, decomposition of organic matter, and production of multiple secondary metabolites like antibiotics (Gopalakrishnan et al. 2020). In addition to these, *Streptomyces* has been frequently reported to be able to fix nitrogen in aerobic condition. The earliest report on nitrogen fixation in the Actinomyces family came from China (Ding et al. 1981). A novel and unusual N_2-fixing system from a chemolithoautotrophic *S. thermoautotrophicus* isolated from a burning charcoal pile was made in 1992 (Gadkari et al. 1992). This nitrogenase was found to be oxygen insensitive and uninhibited by acetylene, ethylene, CO or H_2. Since then, several other reports have suggested the existence of nitrogen fixing *Streptomyces* (Dahal et al. 2017; Sellstedt and Richau 2013). Also, an agar degrading, nitrogen fixing *S. lavendulae* was reported to possess a nitrogen fixing gene *nifU* (Wu et al. 2017). However, the existence diazotrophy in the genus *Streptomyces* is still not confirmed because of the absence of concrete biochemical and genetic evidence. A rigorous study of 561 complete genomes of *Streptomyces* available in NCBI genome database suggest the absence of *nif* genes or its homologus in genus *Streptomyces* (Chapter2). Also, the original report on the existence of nitrogenase in *S. thermoautotrophicus* has been refuted by a group of researchers, who tested the strain in different laboratory settings for its nitrogen fixing ability (MacKellar et al. 2016).

Strain 321G produced ethylene from acetylene in only one replicate out of three, giving inconclusive results for ARA. ARA has been found to give variable results, especially with pure cultures, because nitrogen fixing ability of diazotrophs depends on multiple factors like O_2 concentration, incubation time, temperature and inoculation medium (Estrada-de Los Santos et al. 2018; Hatayama et al. 2005; Khadem et al. 2010; Mirza and Rodrigues 2012). The $^{15}N_2$ assimilation assay showed no evidence of assimilation, and no *nif* genes or homologues were found in the complete genome of strain 321G. This suggested the absence of a traditional nitrogenase enzyme in this strain. However, the ability of this strain to grow on nitrogen free medium, suggests existence of a novel mechanism to procure the nitrogen required for biomass generation. The NFM used was quantitively analyzed to have very little nitrogen, for insufficient to sustain growth of bacteria (chapter 2). In a recent study of dizotrophic *Lactococcus sp.* without *nifHDKENB* genes, a novel pathway for BNF involving genes related to mucilage carbohydrate catabolism, glycan-mediated host adhesin, iron/siderophore utilization, and oxidation/reduction control has been proposed (Higdon, Huang, et al. 2020).

The complete genome sequencing of strain 321G yielded a large genome assembly with 18 contigs. This is a large genome compared to the average size of around 8-9 Mb of genomes for *Streptomyces* (Nindita et al. 2019). A potential reason for such a large genome could be the lenient quality filtering adopted during the processing of nanopore sequence data to capture all the possible *nif* genes and homologues. *Streptomyces* are well known for the repertoire of secondary metabolites like antibiotics (such as chloramphenicol, cypemycin, grisemycin, neomycin, and bottromycins) and range of bioactive compounds actively involved plant beneficial effects and degradation of

organic macromolecules (Sathya et al. 2016). Strain 321G was also found to contain a range of genes for biosynthesis of secondary metabolites and xenobiotic biodegradation and metabolism. In addition, of the 3718 functionally annotated genes, functions related to amino acid metabolism (800) are the most abundant. Although, no genes related to nitrogen fixation could be identified in the up-regulated genes, several genes related to the nitrogen metabolism were identified including a partially complete assimilatory nitrate reduction pathway.

Analysis of genes differentially expressed between the nitrogen rich, and nitrogen deplete medium revealed that strain 321G can utilize a wide range of nitrogenous compounds like nitrate, urea, poly- and monoamines, cofactors, and nucleotides for growth in nitrogen deplete medium (Table 4.2). One of the most upregulated gene with $\log_2$fold change as high as 10.175 in the nitrogen deplete condition was a nitrate/nitrite transporter. Also, the expression level of these transporters was positively corelated with the length of exposure to the nitrogen deplete medium. Uptake of nitrate/nitrite into the cell is the first step in assimilatory nitrate reduction (González et al. 2006) and presence of an assimilatory nitrate reduction pathway in this strain as shown by the whole genome, indicates, this bacteria is utilizing ANRA for producing ammonia in this nitrogen deplete condition. Species of *Streptomyces* have been shown to be able to use nitrate as the sole source of nitrogen using ANRA to produce ammonia necessary for growth (Feng et al. 2014).

Table 4. 2 Differentially expressed pathways involved in nitrogen metabolism.

Function	Log_2foldchange	
	8 h	24 h
Nitrate Assimilation		
Nitrate/nitrite transporter I	6.99	10.175
Nitrate/nitrite transporter II	2.644	5.36
Urease		
ureG, urease accessory protein	5.8	3.37
ureB, urease subunit beta	5.49	3.92
ureC, urease subunit alpha	4.9	2.97
ureF, urease accessory protein	4.49	-
ureA, urease subunit gamma	3.95	2.75
ureAB, urease subunit gamma and beta	2.69	-
Sensory proteins and glutamine synthase		
glnB, nitrogen regulatory protein P-II	6.71	5.011
glnD, uridylyltranserase	4.84	3.5
glnA, glutamine synthase I	2.6	3.44
glnA, glutamine synthase II	1.518	2.49
glnA, glutamine synthase III	-	1.93
Poly- and monoamine metabolism		
potD: spermidine/putrescine transport system: substrate-binding protein	1.4454	-
potA: spermidine/putrescine transport system ATP-binding protein	2.09	-
APA; basic amino acid/polyamine antiporter, APA family	1.44	1.41
Deamination		
Adenosine deaminase	1.061	3.19
Aminodeoxyfutalosine deaminase	1.172	-
Cytosine/creatine deaminase	2.146	-

Strain 321G has significant overexpression of all the structural components of Urease and some accessory proteins for the function of urease. Urease is a nickel-containing enzyme expressed by many pathogenic, marine and soil bacteria, is composed of two or three different polypeptides (UreA (subunit γ), UreB (subunit β), and UreC(subunit α)) and other accessory proteins (Konieczna et al. 2012). Ureases catalyze the hydrolysis of urea into ammonia and carbon dioxide. Stain 321G had significantly higher expression of

all glnA, glnB, and glnD indicating the cells are sensing the scarcity of nitrogen and expressing higher level of glutamine synthase. GlnB and GlnD make a two-component nitrogen regulatory system and primarily involved in the sensing of cellular ammonia level and active Glutamine synthase when cellular ammonia level is low (Ninfa et al. 2000). Glutamine synthase (glnA) is the central enzyme in cellular nitrogen assimilation which is essential for the synthesis of glutamine from ammonia (Forchhammer 2007).

Another strategy applied by the strain 321G to cope with ammonia shortage is deamination of several amine containing biomolecules including nucleotides, amino acids, and other poly- and monoamine compounds. *Streptomyces* have been shown to be able to assimilate polyamines like putrescine, cadaverine, spermidine, and spermine as sources of nitrogen using several aminotransferase pathways (Krysenko et al. 2021:kryse). Several transporters related to these polyamines were also overexpressed in strain 321G, indicating that this isolate is also able to assimilate polyamines as a source of ammonia. Although ammonia is a preferable source of nitrogen and unlike the observation in other study of the oligotrophic actinomyces *Rhodococcus erythropolis* N9T-4 (Yoshida et al. 2014) where an ammonia transporter (amtB) were found to be upregulated in nitrogen limited growth condition, not a single ammonia transport gene was found to be upregulated in this strain. This could probably be the result of stringent growth condition maintained to eliminate all the source of ammonia empirically possible.

4.5 Conclusions

Considering the evidence from the differential gene expression, although strain 321G showed very strong potential to be a diazotrophic bacteria, it appears to be a very

efficient nitrogen scavenger primarily utilizing, nitrate, urea, and multiple amino acids for its growth in extremely nitrogen deficient medium. However, this evidence is based on only 20% of the genes in strain 321G which could be functionally annotated with current cutting age bioinformatics tools. Several differentially expressed genes were hypothetical proteins without any known protein domains hence the existence of another novel nitrogen fixing pathway cannot ruled out and needs further exploration.

Chapter 5

Diversity and Expression of the Nitrogen Cycling Genes in Sioux Prairie Soil Assessed by Meta-transcriptomic Approach.

5.1 Introduction

Prairies are a typical representative of a semi-arid grassland ecosystem which has sufficient moisture to sustain the growth of herbaceous vegetation like grass and bushes, but not enough for the growth of trees (Risser 1985). In the pristine grassland ecosystems like prairie, soil microorganism play a significant role in cycling of various ecologically important elements including nitrogen (Makhalanyane et al. 2015; Steven et al. 2015). Nitrogen is the main growth limiting nutrient and occurs in multiple oxidation states from -3 (NH_4^+) to +5 (NO_3^-). The diverse forms of nitrogen are inter-converted through microbial nitrogen cycling pathways. In addition to the conversion between intermediate forms, ecosystem of influx and efflux of nitrogen is driven by this microbial nitrogen cycle (Dixon and Kahn 2004b). Thus, the study of nitrogen cycling pathways is essential to understand the dynamics and productivity of grassland ecosystem.

The movement of nitrogen to, within and out of the biosphere (Figure 1.2) occurs in a complex network of redox reactions rather than an orderly cycle starting from 1) nitrogen fixation converting dinitrogen to ammonia, 2) ammonia assimilation converting ammonia to organic nitrogen molecules, 3) deamination yielding ammonia back to soil from organic molecules, 4) nitrification oxidizing ammonia to nitrate via nitrite, 5) denitrification releasing nitrogen back to the atmosphere through various intermediates

(nitrite, nitric oxide, and nitrous oxide), and 6) anammox oxidizing ammonia in anaerobic conditions to dinitrogen (Kuypers, Marchant, and Kartal 2018). Recent spurts in DNA sequencing and improvements in in-vitro culture techniques have elucidated the complexity and horizontal mobility of some of the N-cycling genes between taxa of unrelated phylogeny (Isobe and Ohte 2014). Hence, the classical approach of allocating the microorganisms involved in the nitrogen cycle to one step in the cycle is challenged by several observations where a single bacterium has the metabolic aptitude to carry out multiple reactions (Stein and Klotz 2016). The reduction of dinitrogen to ammonia, the major pathway for the entry of dinitrogen into the biosphere, is generally carried out by diazotrophs which also harbor nitrogenase. Nitrogenase is distributed widely in microorganisms from both Archaea and Bacteria, and occur in three variations: Mo-nitrogenase containing MoFe-co at the active site, V-nitrogenase containing VFe-co and Fe-only nitrogenase containing FeFe-co at the active sit (Addo and Dos Santos 2020). The *nifH* gene coding for one of the structural components of nitrogenase has been extensively used for the study of diversity and activity of diazotrophic species in various environments (Gaby and Buckley 2012a). Several species of methanotrophic archaea are found to express genes coding for enzymes responsible for nitrogen fixation, along with nitrification, and denitrification, suggesting that these microbes, when exposed to a wide range of environments can acquire the necessary functions to sustain and propagate in any environment (Stein and Klotz 2016).

The ammonia fixed by diazotrophs is either assimilated to organic nitrogen by prokaryotes, fungi and plants, or aerobically oxidized to form nitrate by the process called nitrification. For a long time only a certain group of chemolithoautotrophic

Proteobacteria (*Nitrosospira, Nirosomonas,* and *Nitrosococcus*) were considered to be able to perform this transformation (Isobe et al. 2011). However, metagenomic sequencing has identified a novel Archaeal group (phylum *Thaumarchaeota*) able to oxidize ammonia, and several representatives of this have already been isolated from soil and hot springs (Venter et al. 2004). Recently, another group of *Proteobacteria* (Comammox) has been identified which is able to carry out the complete oxidation of ammonia to nitrate using the same chemical reactions as above (van Kessel et al. 2015). Another fate of ammonia could be the anaerobic oxidation to dinitrogen (anammox) which is one of the methods for removal of bioavailable nitrogen back to the atmosphere.

Nitrate can be reduced back to ammonia in two processes. Assimilatory nitrate reduction to ammonia (ANRA) is carried out by diverse eukaryotes, bacteria and archaea by assimilatory nitrate reductases (NAS) (Moreno-Vivián et al. 1999). ANRA, as the name suggests, assimilates nitrogen directly to the biomass of the organism carrying out this process. On the other hand dissimilatory nitrate reduction to ammonia (DNRA) is associated with respiration and returns back nitrogen in the form of ammonia, that can be assimilated by other organism (Pajares and Bohannan 2016). DNRA is also carried out by organisms of all three domains, and the reaction is catalyzed either by membrane bound nitrate reductase (NAR) or periplasmic nitrate reductase (NAP). This reaction is sometimes shared by denitrifying bacteria and a strict demarcation between these two processes may not be possible (Tsementzi et al. 2016). Reduction of nitrite to dinitrogen is called denitrification and occurs via multiple intermediate compounds like nitric oxide and nitrous oxide. Nitrite to nitric oxide reduction is carried out by bacteria such as Proteobacteria and Bacteroidetes with the enzyme NirK or NirS (Maia and Moura 2014).

Nitric oxide reduction to nitrous oxide is carried out by the enzyme NorBC and found in denitrifying bacteria like *Pseudomonas* and some fungi (Shoun et al. 2012). The nitrous oxide is finally reduced to dinitrogen by NosZ which occurs in diverse bacteria from Proteobacteria, Bacteroidetes, and Chlorobi to the archaeal Crenarchaeota and Halobacteria (Cabello, Roldán, and Moreno-Vivián 2004).

The presence and level of expression of nitrogen cycling genes in an ecosystem is usually inferred by the presence and expression of marker genes but these methods are often plagued by the suboptimal PCR amplification and primer bias, and fail to capture the complete active microbial community due largely to lack of fully conserved region in these genes pools (Albright et al. 2019). Therefore, this chapter explores the application of high throughput sequencing (HTS) of a complete soil transcriptome to identify the major nitrogen cycling genes and diversity of microorganism involved in the nitrogen cycle.

5.2 Materials and methods

5.2.1 Sampling site

Samples were collected from the Sioux Prairie conservation area located on the north of Colman, SD. This is a natural grassland, conserved by the Nature Conservancy (TNC). The field is located at 231st St, Colman, SD 57017 (Figure 5.1).

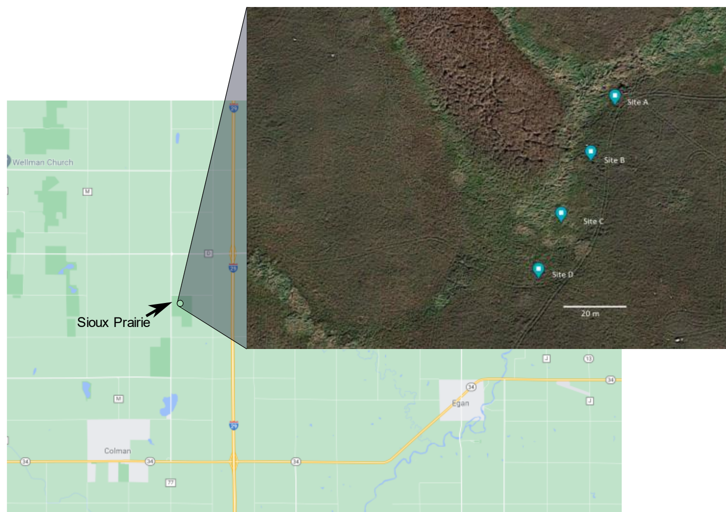


Figure 5. 1 The location of sampling site at the Nature Conservancy (TNC) site at Sioux Prairie. 4 sites 20 meters apart were selected for collecting the samples. Source google.map

Samples were collected from four sites (Site 1, Site2, Site 3 and Site 4) which were located 20 meters apart at the center of the prairie (44.0356880, -96.7847880; 44.0355300, -96.7848840; 44.0353580, -96.7849990; 44.0352010, -96.7851960). Soil cores of 10 cm depth and 2.5 cm in diameter were collected. A LifeGuard™ Soil Preservation Solution (Qiagen) was added to protect RNA and DNA degradation during the transfer to the laboratory. Soil samples were collected in sterile and airtight vials and stored at 4°C when transferring to the laboratory for further processing.

5.2.2 Extraction of total RNA

The RNeasy PowerSoil Total RNA Kit from Qiagen was used for isolation of total RNA. Soil samples were mixed, and 2 g of sample was added to a15ml PowerBead

Tube, and total RNA extracted according to the product's protocol. The extracted RNA was quantified using fluorometry using Qubit 3 fluorometer and stored at -80ºC until further processing.

5.2.3 RNA Quality check and HTS

The RNA samples were outsourced to Novogene Corporation Inc. for further quality check and sequencing. The quality of RNA was assessed by resolving in 1 % Agarose gel electrophoresis at 180 V for 16 minutes. RNA integrity was measure on an Agilent 2100 and once the RNA samples passed the quality requirements, rRNA was removed using the Ribo-Zero kit. cDNA was synthesized from the purified mRNA using random hexamer primers. The libraries were subjected to paired end sequencing on a Illumina sequencer.

5.2.4 Quality check and processing of sequence data

The sequences obtained from Novogene were subjected to quality check using FastQC (Li et al. 2015). The taxonomy of reads was identified with Kaiju (Menzel, Ng, and Krogh 2016) using the prokaryotic database as reference. The sequences were assembled using Trinity (Grabherr et al. 2011). The assembled transcripts were also classified using Kaiju. The reads were aligned against the assembled transcripts and a count of mapping was generated by using Salmon (Patro et al. 2017). The transcripts were identified by using a Blastx search against the NCBI nr protein database, HMMER search for Pfam, and Blastp search for genes predicted from assembled transcripts. All the results from Blastx, Blastp, HMMER were compiled to the SQLlite database using the trinitiy boilerplate templet. Trinity boiler plate comes with precompiled GO, KEGG

and taxonomy databases. The N-cycling pathways were identified with the Gene ontology (GO) terms listed in Table 5.1.

Table 5. 1 GO terms used in this study to identify the transcripts involved in nitrogen cycling pathways.

SN	GO term	N cycling pathway
1	GO:0016163	Nitrogenase activity
	GO:0009399	Nitrogen fixation
2	GO:0006542	Glutamine biosynthesis process
3	GO:0019239	Deaminase activity
4	GO:0018597	Ammonia monooxygenase
5	GO:0047991	Hydroxylamine dehydrogenase
6	GO:0019332	Nitrite oxidoreductase
7	GO:0042128	Nitrate assimilation
8	GO:0019333	Denitrification
9	GO:0044222	Hydrazine synthase

5.3 Results

5.3.1 Quantification and quality of RNA extract

On average 2.55 μg of total RNA was extracted from each site. The 1% agarose gel electrophoresis showed the clear 23S and 16S rRNA bands indicating the RNA sample was intact and the Agilent RNA integrity number ranged from 7.4 to 7.6, indicating good quality of the RNA extract (Table 5.2).

Table 5. 2 Quantification and quality of RNA extracted.

Sample Name	Concentration (ng/ul)	Integrity value*	Gel image
A1R10	45	7.6	
A2R10	45	7.4	
A3R10	122	7.7	
A4R10	64	7.6	

*RNA integrity categories 1 (totally degraded) -10 (intact).

5.3.2 Sequencing quality

Total mRNA extracts were sequenced on an Illumina platform and 212.4 million paired end mRNA sequences were obtained in total from the four soil samples. The reads were paired and 150 bp long on average. The phred quality score for each position was at least 35 indicating good sequencing quality. The GC content on average was 62.48%, indicating the abundance of organisms with high GC ratio in these soil samples (Table 5.3).

Table 5. 3 Quality of the sequencing reads.

Sample	Raw reads	Raw data (Gb)	Effective (%)	Error (%)	Q20 (%)	Q30 (%)	GC (%)
A1R10	49068766	7.4	98.45	0.02	98.12	94.81	63.22
A2R10	50606062	7.6	98.28	0.02	98.04	94.61	61.72
A3R10	51510666	7.7	98.91	0.02	97.96	94.47	62.87
A4R10	61219586	9.2	98.14	0.02	98.1	94.74	62.11

Raw reads= total number of reads (both ends) in each sample.
Raw data= Total number of bp sequenced in each sample represented in Giga base
Effective = percentage of clean reads among total reads
Error = base error rate
Q20, Q30 = corresponds to the phred score of 20 and 30. GC: GC content

5.3.3 De novo assembly of transcriptome

All the paired end reads were assembled into 6198722 transcripts using the de novo assembler, Trinity. The average length of transcripts was found to be 341.92. These transcripts were not uniformly distributed across the four samples (Table 5.4). The coverage of the assembled transcripts in the samples was found to be 76.13 %, 71.61%, 72.36%, and 84.49% in A1R10, A2R10, A3R10, and A4R10 respectively. Only 54.38% of the transcripts were identified by finding the best hit on the NCBI nr protein database using Blastx. A total of 5263238 genes were predicted from the total transcripts based on the presence of the open reading frames, and 34.9 % of these genes were identified by an HMMER search for protein domains on Pfam.

Table 5. 4 Trinity Assembly summary

	Number of Reads	**Median length**	**Average length**	**Total assembled bases**	**N50**
Transcripts	6198722	272	341.92	2119456127	326
Genes	5263238	268	324.33	1707002035	303

N50 is the mean number of base pairs in all transcripts that, ordered by length, make up 50% of the assembly.

5.3.4 Diversity of the prokaryotes

RNA from bacteria, archaea and viruses was obtained in all the samples, with bacteria being the most dominant (~49%) (Figure 5.2). Archaea and bacteria represented ~ 1% of the total reads and ~50% of the reads could not be classified as prokaryotes. The transcripts in the assemblies also followed a similar pattern with some variations. Less than 1% of transcripts were classified as archaea and viruses. The bulk of transcripts (67.5%) were classified as bacteria and only 31.76% of the transcripts could not be classified as any prokaryotes.

On analysis of the classified reads and transcripts at the phylum level, the most abundant phyla were Proteobacteria (15-22%), Actinobacteria (9-18%), Acidobacteria (4.6-7%), and Chloroflexi (1-2%). These were followed by Planctomycetes, Candidatus Rokubacteria, Verrucomicrobia, Bacteroidetes, and Firmicutes each consisting of ~1% of total reads.

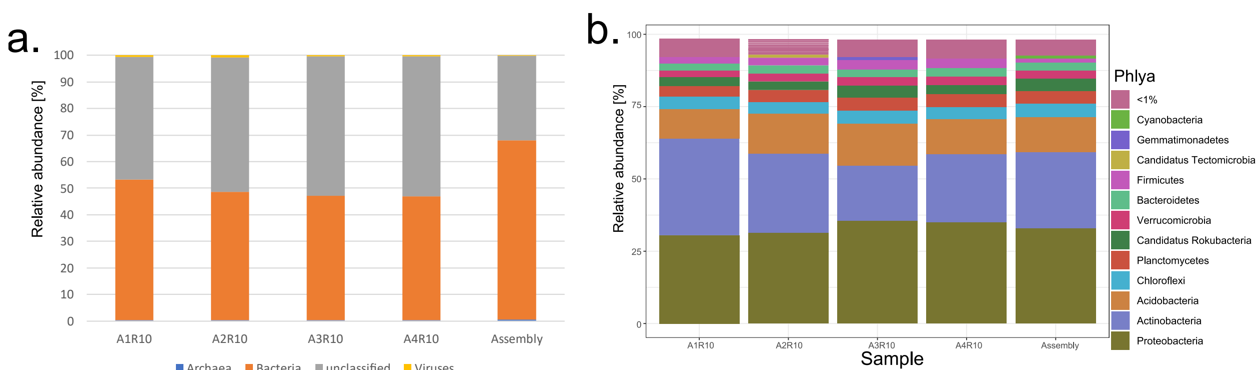


Figure 5. 2 Diversity of prokaryotes based on the total RNA reads and the assembled transcripts. Diversity of the prokaryotic super-kingdoms (a) and the prokaryotic phyla excluding the unclassified reads (b).

5.3.5 N_2 cycle

Diversity and expression of the nitrogen cycling genes involved in nitrogen fixation, ammonia assimilation, mineralization, nitrification, nitrate reduction, and denitrification pathways were successfully identified using the GO terms.

Nitrogen fixation

A very small number of transcripts (24) out of total 6.2 M transcripts were found to code for the components of nitrogenase enzyme (NifDKH ENB or Vnf/AnfG). *nifH* (10 transcripts) was the most abundant of the nitrogenase proteins followed by *nifDK* (5 transcripts each). Two transcripts each of *nifE* and *nifN* were detected and *nifB* was not

detected at all. Organisms expressing *nif* genes belonged to phylum Proteobacteria, Cyanobacteria, Chloroflexi, and Chlorobi and archaeal phylum Euryarchaeota. Alphaproteobacteria were most common and represented by *Bradyrhizobium, Azospirillum, Mesorhizobium,* and *Zymomonas.* Gammaproteobacteria were represented by *Azotobacter, Teredinibacter,* and *Klebsiella. Herbaspirillum* was the only Betaproteobacteria expressing *nif* genes. In addition to these Proteobacterial classes, *Acidithiobacillus* from class Acidithiobacilli was also found to express *nifH* genes. Firmicutes, Actinobacteria, Chlorobi, Cyanobacteria, and Chloroflexi were all represented by the single genera *Clostridium, Frankia, Chlorobaculum, Nostoc,* and *Rosiflexi* respectively. Euryarchaeota were represented by *Methanomicrobia and Methanobacteria* (Figure 5.3 a).

The expression of the total nitrogenase related genes was found to be in the range of 0 to 7.2 transcripts per million, with site 1 showing the least expression and site 3 showing the highest expression. The expression of *nifH* was found in the range of 0 to 1.96 transcripts per million (Figure 5.3b). Although a very small number of structural components of nitrogenase were detected from multiple taxa, 16,270 transcripts related to regulation of expression of nitrogenase were detected, suggesting the limited expression of nitrogenase could be the result of precise regulatory mechanism repressing the expression of *nif* genes in these grassland populations.

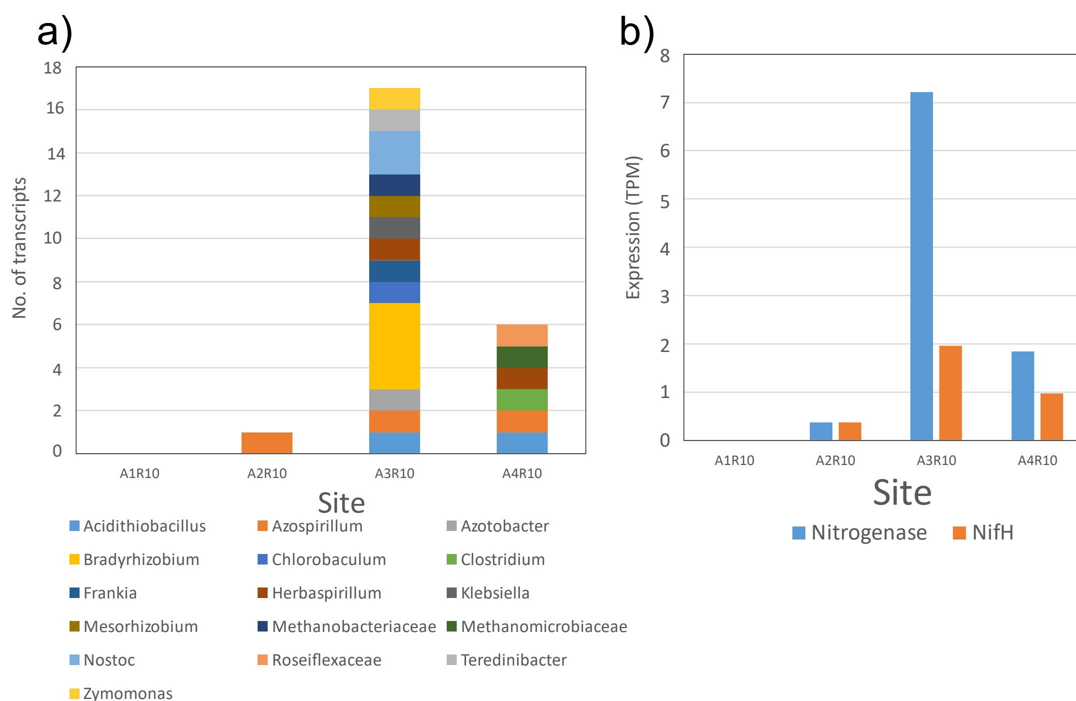


Figure 5. 3 Diversity (a) and expression (b) of transcripts encoding structural nitrogenase components (nifHDKENB)

Nitrification

Ammonia can be utilized by certain chemolithoautotrophs as the source of electrons and energy, oxidized to nitrate by two groups of bacteria which carry out each half of the pathway. In this study genes expressing ammonia monooxygenase and hydroxyamine dehydrogenase were detected from the single genus *Nitrosomonas*. The expression of *AmoA* was in the range of 0.2 to 1.4 TPM, with site 1 showing the highest expression. The expression of *Hao* ranged from 0.06 to 3.1 TPM, with site 3 showing the

highest expression. These expression levels are even less than the expression of nitrogenase genes (Figure 5.4).

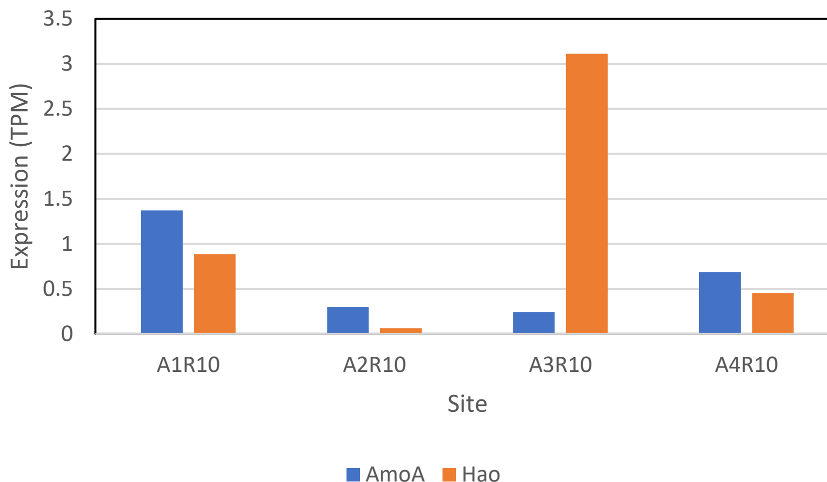


Figure 5. 4 Expression of the nitrification genes across the four sites. Oxidation of ammonia to nitrate occurs in two steps. Genes necessary for the conversion of ammonia to nitrite (AmoA and Hao) were detected but nitrite oxidation to nitrate was not detected.

Ammonia assimilation and mineralization

The first step towards utilization of ammonia by non-diazotrophic organisms is the synthesis of glutamine by glutamine synthase (*glnA*), after which nitrogen is incorporated into various cellular macromolecules. Ultimately, the nitrogen is converted back to ammonia by deamination during the decomposition of dead organic matter. Both glutamine synthesis and deamination reactions do not change the oxidation state of ammonia. In this study, both eukaryotic and prokaryotic organisms were found capable of these transformations (Figure 5.5). A higher number of transcripts involved in assimilation was recorded compared to deamination. Assimilation was dominated by Actinobacteria, Proteobacteria, and Cyanobacteria. Actinobacteria producing *glnA* were *Streptomyces, Frankia,* and *Mycobacterium* suggesting that these are efficient ammonia utilizing species among grassland soil Actinobacteria. Among the proteobacteria, species

of Alphaproteobacteria, Betaproteobacteria, and Gammaproteobacteria expressed *glnA*. Alphaproteobacteria were represented by *Azospirillium, Azorhizobium, Bradyrhizobium, Rhizobium, Sinorhizobium, Rhodobacter.* Similarly, Betaproteobacteria were represented by *Cupriavidus, Neisseria,* and *Methyloversatilis,* and Gammaproteobacteria were represented by *Pseudomonas, Acinetobacteria, Azotobacter,* and *Methylococcus.* Cyanobacteria were represented by *Nostoc* and *Synechococcus* and Firmicutes were represented by *Bacillus, Lactobacillus, Staphylococcus,* and *Clostridium.*

On the other hand, deamination was dominated by Euryarchaeota, Proteobacteria, and fungi from Eukarya. Euryarchaeota represented by *Pyrococcus* and *Saccharolobus.* Proteobacteria represented by Gamma (*Pseudomonas, Escherichia, Salmonella, Shigella, Haemophilus*) and Epsilon (*Helicobacter*). The ammonia assimilating and mineralizing organisms were uniformly distributed across the four sites with some minor differences.

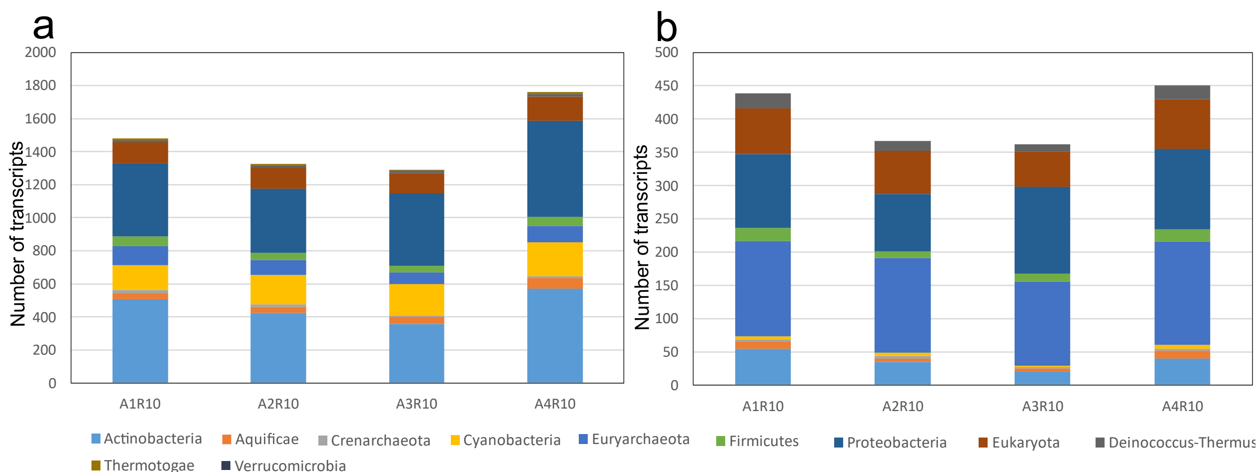


Figure 5. 5 Phylogenetic association of the ammonia assimilating gene glnA(a) and the ammoina mineralizing gene deaminase (b).

The expression of *glnA* and deaminase was significantly different, with *glnA* expressed much higher with average expression of 271.2 TPM and deaminase only 61.175 TPM. This indicates that ammonia is being used by the organisms at much higher rate than it is being cycled back to the soil (Figure 5.6).

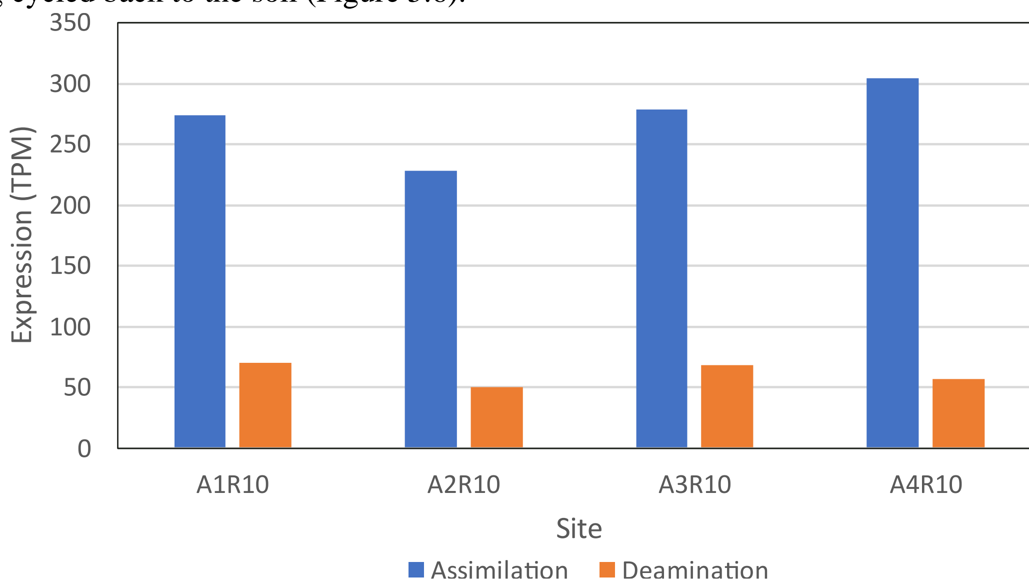


Figure 5. 6 The measurement of expression of assimilation (glnA) and Mineralization (deaminase) across the four sites. The expression of glnA is found to be much higher than that of deaminase.

Nitrate assimilation

Nitrate is the most oxidized form of nitrogen and can either be reduced back to ammonia or can be reduced to dinitrogen, thereby making it inaccessible to most cells. The process by which nitrate is reduced to ammonia is broadly termed nitrate assimilation. It can occur via two pathways: Assimilatory nitrate reduction to ammonia (ANRA) and Dissimilatory nitrate reduction to ammonia (DNRA). Unlike denitrification, nitrogen is cycled back to the biosphere in these processes. The first step in nitrate assimilation is reduction of nitrate to nitrite which is carried out by different enzymes in different organisms. In this study *NR*, *narB, nasA, narG, and napA* were used as the marker genes to study expression of enzymes reducing nitrate to nitrite (Figure 5.7).

Nitrate reductase was the most abundant N-related transcript, and expressed mainly by eukaryotes (Algae, fungi, and phytoplankton). *narG* was the next most abundant transcript, expressed by four genera: *Mycobacterium* (Actinobacteria), *Escherichia* (Proteobacteria), *Bacillus* (Firmicutes), and *Haloferax* (Archaea). It was followed by *napA*, which is expressed by diverse bacteria from Proteobacteria and Firmicutes. Among the bacteria expressing *napA*, *Symbiobacterium* (Firmicutes) was most abundant followed by Proteobacteria (*Bradyrhizobium, Cupriavidus, Dechloromonas, Desulfovibrio, Rhodopseudomonas, Rhizobium,* etc). It was followed by *narB* which was expressed mainly by Archaea (*Haloferax*) and Cyanobacteria (*Synechococcus* and *Synechocystis*). This was followed by *nasA* expressed by *Klebsiella* and *Haloferax.* ANRA is catalyzed by *nirA* (expressed by *Haloferax* and *Phormidium*) and NIT-6 (expressed by *Bacillus* and some eukaryotes). DNRA is catalyzed by *nirB* (expressed by *Escherichia* and *Klebsiella*).

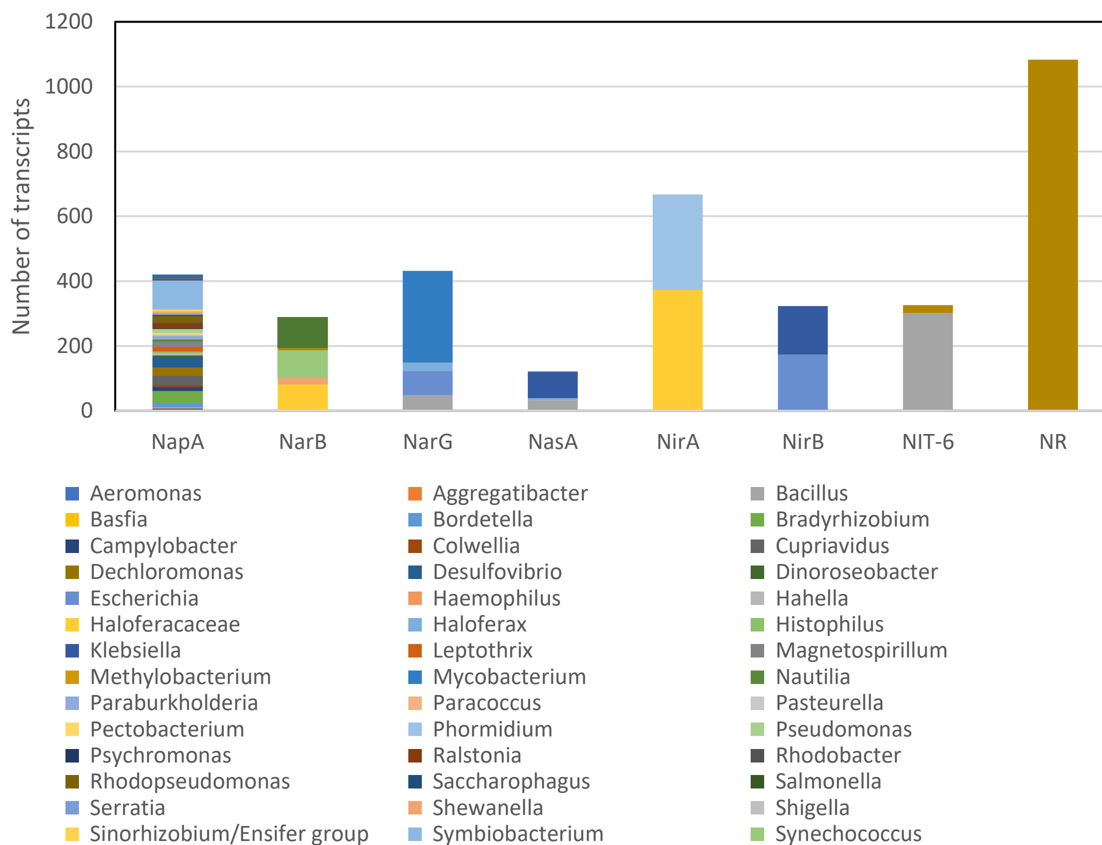


Figure 5. 7 Diversity of genes involved in multiple pathways of nitrate assimilation.

The expression level of nitrate reduction genes follows similar pattern to the number of transcripts found in each gene. For the enzymes involved in reduction of nitrate to nitrite, NR was the most abundant followed by *narG, napA, narB and nasA*. For the ANRA *nirA* was most expressed followed by NIT-6 and for DNRA only *nirB* was expressed (Figure 5.8).

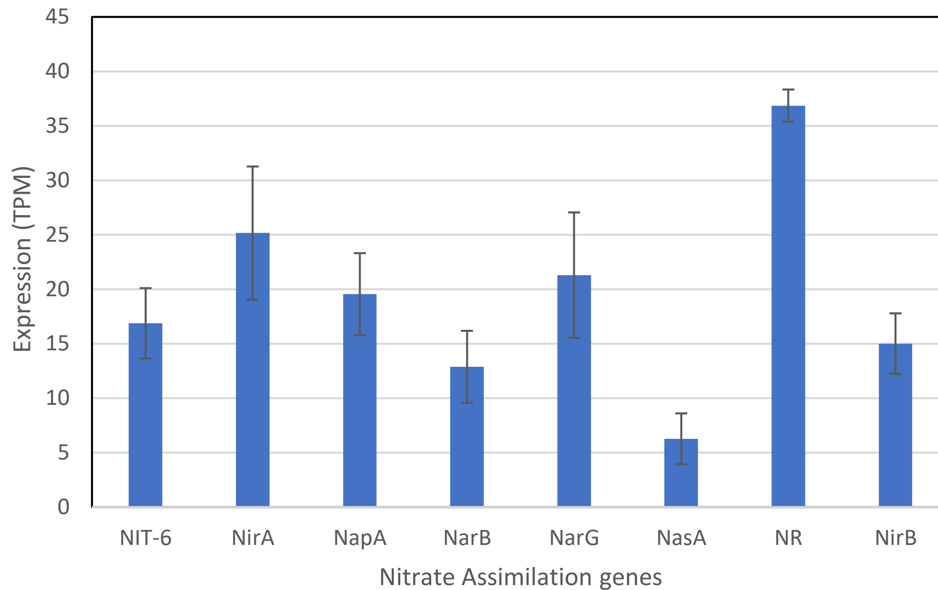


Figure 5. 8 Expression of nitrate assimilation genes.

Denitrification

It is the process by which bioavailable nitrogen is converted to dinitrogen. It occurs either by the reduction of nitrate to dinitrogen through multiple intermediates ($NO_3^- > NO_2^- > NO > N_2O > N_2$), or by direct oxidation of ammonia by anammox. Marker genes *nirK, norB, nosZ, nrfH* are involved in reduction of nitrate and Hzs is involved in direct oxidation of ammonium. In this study, *hzs* was the most abundant and highly expressed transcript which aligned to the single bacteria *Candidatus* Kuenenia (Planctomycetes). *nosZ* was the next most abundant transcript expressed by *Bradyrhizobium, Pseudomonas, Brucella, Sinorhizobium, Cupriavidus,* and *Achromobacter*. It was followed by *nirK* which was also expressed mainly by Proteobacteria (*Alkaligenes, Pseudomonas, Rhizobium, Rhodobacter,* and

Sinorhizobium). It was followed by *norB* (*Pseudomonas*) and *nrfH* (*Wolinella).* The expressions of denitrifying genes were also corelated with the number of the transcripts recorded with some variations among different sites (Figure 5.9).

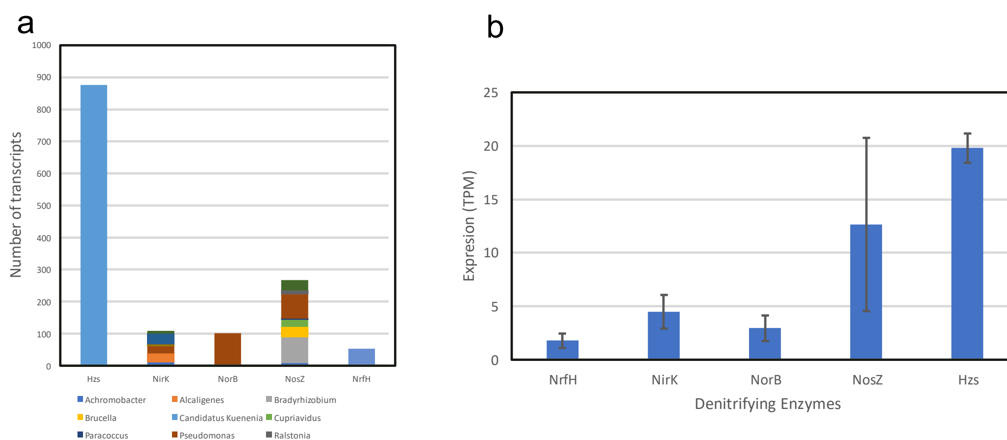


Figure 5. 9 Diversity (a) and expression (b) of marker genes involved in the denitrification process.

Expression levels of N_2 cycling reactions

Expression levels of various pathways involved in nitrogen cycling were evaluated by selecting a marker gene for each pathway. Expression of these marker gene was detected in all the pathways except nitrite oxidation (a part of nitrification) and hydrazine dehydrogenase (a part of anammox). The ammonia monooxygenase was the least expressed pathway and nitrogen fixation the second least expressed pathway. The most expressed pathway is ammonia assimilation by *glnA*, followed by mineralization. All the other internal cycling pathways showed intermediate expression levels (Figure 5.10).

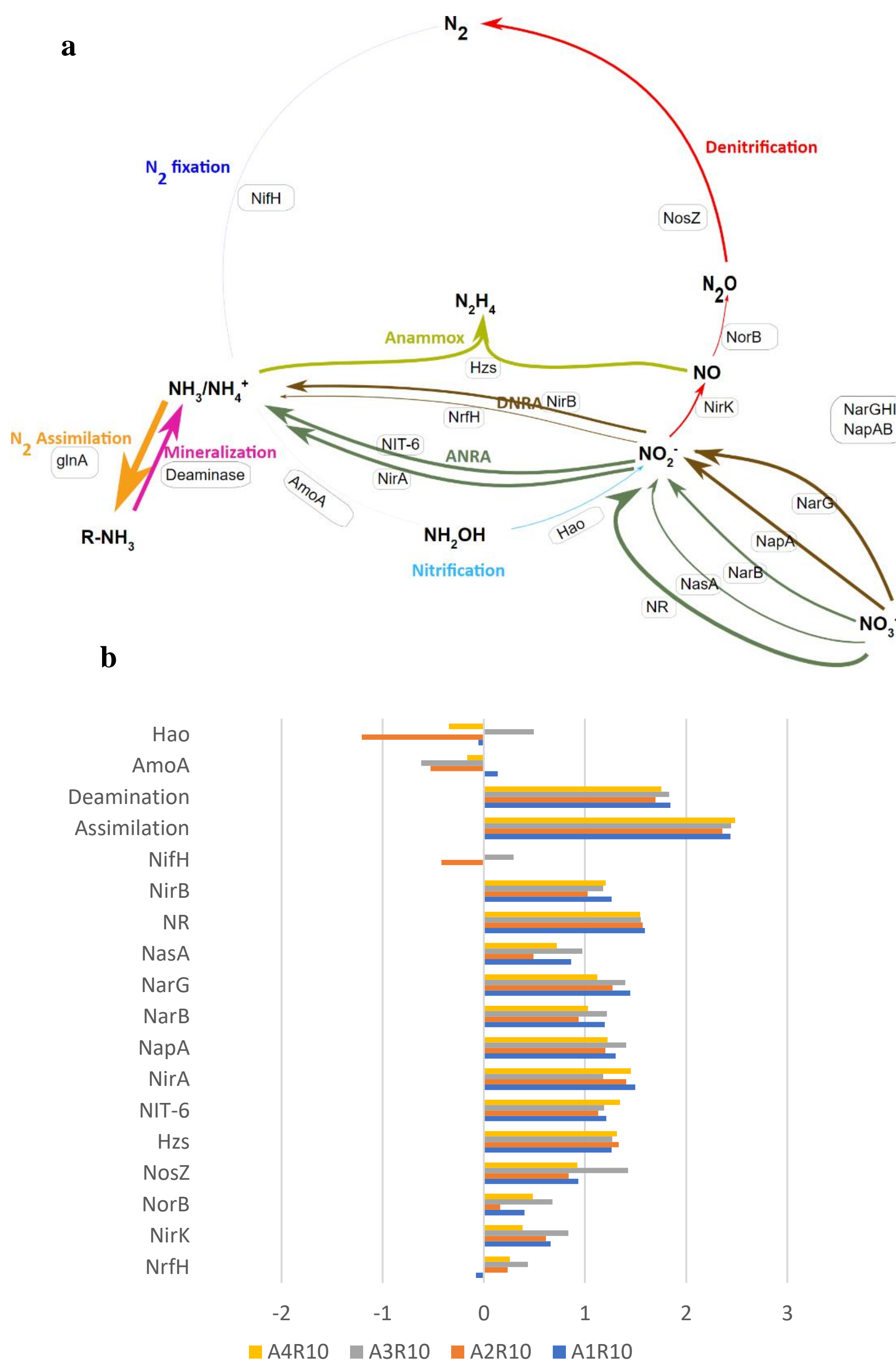


Figure 5. 10 Nitrogen cycle showing log_2 expression of marker genes for each reaction as indicated by the thickness of the line (a). The expression level of N-cycling marker genese shown as log_{10} of average expresion of genes (b).

5.4 Discussion

The primary goal of this study was to quantify the expression levels of multiple pathways in the nitrogen cycle and explore the diversity of organisms actively involved in nitrogen cycling in a prairie soil environment. I used an approach that is free from the limitations inherent to contemporary PCR-based methods widely used for similar studies. The NiCE Chip is one of the cutting-edge techniques where multiple functional genes can be quantified in a single run using qPCR and microfluidic techniques. Multiple studies have successfully used the NiCE chip to quantify multiple functional nitrogen cycle genes, all at the same time and shown the results to be comparable to individual qPCR amplifications (Oshiki, Segawa, and Ishii 2018; Spurgeon, Jones, and Ramakrishnan 2008). However, this technique fundamentally relies on PCR and primer sets that must amplify all alleles of the gene, risking biased results. Similarly, other PCR independent methods like NanoString can eliminate the primer bias but require the development of N cycle specific mRNA probes (Geiss et al. 2008). Stable Isotope probing to identify members involved in active metabolism can be biased because of potential leakage of isotopic N to non-diazotrophic or non N cycling organisms (Bell et al. 2011). Furthermore efficient separation of ^{15}N-labelled, nucleic acid is challenging (Neufeld et al. 2007). In addition, all these methods rely on high through put sequencing (HTS) for elucidating the diversity of the active N-cycling species (Cloutier et al. 2019). This study successfully shows that in-depth meta-transcriptomic sequencing of total mRNA can give both the diversity of N-cycling organisms and the expression level of multiple pathways in N cycle without the weakness of imperfect primers. In addition to this, complete meta-

transcriptomic data also provides information on many other metabolic pathways actively occurring in the given sample (Moran et al. 2013).

Meta-genomic sequencing has been used extensively to study the diversity of N-cycling organism (Albright et al. 2019; Chen et al. 2015; Su et al. 2012). In a comparative study where 365 soil metagenomes from all around the globe were studied across the eight nitrogen cycling pathways, N-cycling pathways were identified in Bacteria, Archaea, and Fungi (Nelson, Martiny, and Martiny 2016). In this study also, the most dominant organisms harboring N-cycling genes were Bacteria, followed by Archaea and other eukaryotes like Fungi. Of the 3.37 M annotated transcripts, 12,603 transcripts (0.374%) were found to code for at least one of the N-cycling marker genes used. This observation is comparable to the 0.5% of annotated sequences in a soil metagenome samples being associated with one of the eight N pathways reported by (Nelson et al. 2016). Compared to the metagenomic approach, a transcriptomic approach provides a reflection of the active microorganism, and the genes they express, that are directly related to the metabolic functions of interest. Recently, several other studies have also employed HTS of complete transcriptomes to study N-cycling genes (Jewell et al. 2016; Ren et al. 2018; Xiao et al. 2019).

Among the total reads, the most abundant phyla were Proteobacteria (15-22%), Actinobacteria (9-18%), Acidobacteria (4.6-7%), and Chloroflexi (1-2%), followed by Planctomycetes, Candidatus Rokubacteria, Verrucomicrobia, Bacteroidetes, and Firmicutes, each consisting of ~1% of the total reads. This result is almost identical to the diversity obtained using HTS of V3-V4 region of 16S rRNA (Das et al., Unpublished

data). Verrucomicrobia were found to be slightly higher in abundance by 16S rRNA indicating, the species from this phylum are metabolically less active than other phyla in this soil environment. The diversity of soil microorganisms was shown to be significantly affected by the physio-chemical properties of the soil (Ausec, Kraigher, and Mandic-Mulec 2009; Hayden et al. 2010; Makhalanyane et al. 2015; Saleh-Lakha et al. 2005). For example, a metatranscriptomic study from a saline soil (Ren et al. 2018) showed only Proteobacteria, Actinobacteria, Firmicutes and Euryarchaeota to be dominant phyla.

Among the nitrogen cycling pathways, nitrogen fixation genes were expressed only in the phyla Proteobacteria, Cyanobacteria, Chloroflexi, Chlorobi, and the archaeal phylum Euryarchaeota. No evidence of alternative nitrogenase was detected and the expression level for the marker gene *nifH* was found to be only 0.83 TPM. This indicates very low rate of nitrogen fixation occurring in the prairie soil. This result is in congruence with the result from NiCE chip data from the same sample (Das et al., unpublished data). Similar low expression of *nifH* was observed in another meta transcriptomic study in saline soil (Ren et al. 2018). This observation is also comparable with the results from metagenomic data, where N fixation was detected only 4.6 times per million sequences (Nelson et al. 2016). Meta-transcriptomes are expected to yield fewer genes than meta-genomic data as they are based only on the genes being actively transcribed. Although the structural components of nitrogenase were expressed in low number, significantly higher number of nitrogen fixation regulatory transcripts (16,270) were expressed in this study, indicating the tight regulation of energy expensive nitrogenase.

Nitrification is another least expressed N-cycling pathway (6.1 times per million sequence (Nelson et al. 2016)) and expressed by a very limited number of microorganisms, suggesting lack of horizontal gene transfer (Albright et al. 2019). They were the least expressed N-cycling pathway in this study. *amoA* had the expression of only 0.6 TPM and Hao had 1.1 TPM. Recently, various Archaea and proteobacteria are shown to be capable of nitrification (Hayatsu, Tago, and Saito 2008) but in this study only *Nitrosomonas* was found to express nitrification genes. Ammonia assimilation is almost ubiquitously (93 % of genomes) in bacteria (Albright et al. 2019) and in this study also it is the most expressed N-cycling pathway and expressed by diverse microorganism from both prokaryotes and eukaryotes. In this study the most abundant goup expressing ammonia assimilation were found to be Actinobacteria like *Streptomyces, Frankia,* and *Mycobacterium* which are also one of the most common bacteria isolated in nitrogen deficient medium without the evidence of nitrogenase (Chapter 3). This high expression of *glnA* could be the potential reason for this characteristic. Multiple pathways know to be involved in nitrate reduction were observed in this study. The diversity of organisms expressing these genes is very much comparable to that observed in other studies (Kuypers et al. 2018; Ren et al. 2018) . The anammox pathway was only observed in one taxon *Candidatus* Kuenenia which is a Planctomycetes and one of the most studied bacteria for the anammox pathway (Speth et al. 2012).

The expression level of N-cycling gene as a whole was found to be very comparable to other studies (Ren et al. 2018) (Das et al., Unpublished). One of the common features reported in all these studies is the low rate of nitrogen fixation compared to higher denitrification. These approaches using marker genes to study several

pathways are found to yield false positive results because of presence of incomplete functional pathways (Albright et al. 2019). To eliminate this possibility, thorough examination of each pathway with all the necessary auxiliary genes may be required. Also, according to the nitrogen-metabolism network view proposed by Kuypers et al.,2018, irrespective of the microbial composition the ecosystems and biosphere are always in nitrogen homeostasis, where most of the nitrogen is cycled within the ecosystem and very little is lost back to dinitrogen and this lost nitrogen is always balanced by nitrogen fixation. Hence, the approach to quantify nitrogen pathways may provide quick way to get the estimate of the nitrogen cycle, its results should always be taken with a grain of salt.

5.5 Conclusion

Meta-transcriptomic sequencing which is free of biases observed in PCR based methods was successfully implemented to characterize the diversity and expression of one of the most important ecological pathways by which nitrogen is cycled via multiple intermediates in different life forms. Study of the multiple structural genes of nitrogenase showed that nitrogen fixation occurs in extremely low level compared to other pathways in nitrogen cycle in the prairie soil. Similarly, nitrification also occurs in very low level and is carried out by single taxon, *Nitrosomonas* and nitrate reduction is the most expressed and carried out by most diverse organisms. Bias observed with similar other studies is related to false positive results because of use of single marker genes, as the expression of the whole pathway. Meta-transcriptomic approach might provide solution

to this problem as well if the sequences could be binned in similar approach as the one used in metagenomic approach.

References

Abadi, V. A. J. M., M. Sepehri, H. A. Rahmani, H. K. Dolatabad, M. Shamshiripour, and B. Khatabi. 2021. "Diversity and Abundance of Culturable Nitrogen-fixing Bacteria in the Phyllosphere of Maize." *Journal of Applied Microbiology* 131(2):898–912. doi: 10.1111/jam.14975.

Abayasekara, Lalanika M., Jennifer Perera, Vishvanath Chandrasekharan, Vaz S. Gnanam, Nisala A. Udunuwara, Dileepa S. Liyanage, Nuwani E. Bulathsinhala, Subhashanie Adikary, Janith V. S. Aluthmuhandiram, Chrishanthi S. Thanaseelan, D. Portia Tharmakulasingam, Tharaga Karunakaran, and Janahan Ilango. 2017. "Detection of Bacterial Pathogens from Clinical Specimens Using Conventional Microbial Culture and 16S Metagenomics: A Comparative Study." *BMC Infectious Diseases* 17(1):631. doi: 10.1186/s12879-017-2727-8.

Achouak, Wafa, Philippe Normand, and ThierryYR 1999 Heulin. 1999. "Comparative Phylogeny of Rrs and NifH Genes in the Bacillaceae." *International Journal of Systematic and Evolutionary Microbiology* 49(3):961–67. doi: 10.1099/00207713-49-3-961.

Adam, Panagiotis S., Guillaume Borrel, Céline Brochier-Armanet, and Simonetta Gribaldo. 2017. "The Growing Tree of Archaea: New Perspectives on Their Diversity, Evolution and Ecology." *The ISME Journal* 11(11):2407–25. doi: 10.1038/ismej.2017.122.

Addo, Maame A., and Patricia C. Dos Santos. 2020. "Distribution of Nitrogen-Fixation Genes in Prokaryotes Containing Alternative Nitrogenases." *ChemBioChem* 21(12):1749–59. doi: 10.1002/cbic.202000022.

Agafonova, N. V., N. V. Doronina, E. N. Kaparullina, D. N. Fedorov, A. B. Gafarov, O. I. Sazonova, S. L. Sokolov, and Yu. A. Trotsenko. 2017. "A Novel Delftia Plant Symbiont Capable of Autotrophic Methylotrophy." *Microbiology* 86(1):96–105. doi: 10.1134/S0026261717010039.

Ahmad, Farah, Iqbal Ahmad, Farrukh Aqil, Aijaz Ahmed Wani, and Yogesh S. Sousche. 2006. "Plant Growth Promoting Potential of Free-Living Diazotrophs and Other Rhizobacteria Isolated from Northern Indian Soil." *Biotechnology Journal* 1(10):1112–23. doi: 10.1002/biot.200600132.

Alarico, Susana, Mafalda Costa, Marta S. Sousa, Ana Maranha, Eva C. Lourenço, Tiago Q. Faria, M. Rita Ventura, and Nuno Empadinhas. 2014. "Mycobacterium Hassiacum Recovers from Nitrogen Starvation with Up-Regulation of a Novel Glucosylglycerate Hydrolase and Depletion of the Accumulated Glucosylglycerate." *Scientific Reports* 4(1):6766. doi: 10.1038/srep06766.

Albright, Michaeline B. N., Bibek Timalsina, Jennifer B. H. Martiny, and John Dunbar. 2019. "Comparative Genomics of Nitrogen Cycling Pathways in Bacteria and Archaea." *Microbial Ecology* 77(3):597–606. doi: 10.1007/s00248-018-1239-4.

Andrew Jasniewski, Nathaniel Sickerman, Yilin Hu, and Markus Ribbe. 2018. "The Fe Protein: An Unsung Hero of Nitrogenase." *Inorganics* 6(1):25. doi: 10.3390/inorganics6010025.

Anisimova, Maria, Manuel Gil, Jean-François Dufayard, Christophe Dessimoz, and Olivier Gascuel. 2011. "Survey of Branch Support Methods Demonstrates Accuracy, Power, and Robustness of Fast Likelihood-Based Approximation Schemes." *Systematic Biology* 60(5):685–99. doi: 10.1093/sysbio/syr041.

Atlas, Ronald M. 2010. *Handbook of Microbiological Media*. 4th ed. Boca Raton: CRC Press.

Ausec, Luka, Barbara Kraigher, and Ines Mandic-Mulec. 2009. "Differences in the Activity and Bacterial Community Structure of Drained Grassland and Forest Peat Soils." *Soil Biology and Biochemistry* 41(9):1874–81. doi: 10.1016/j.soilbio.2009.06.010.

Ausubel, F. M. 1988. *Current Protocols in Molecular Biology*. 1988th ed. John Wiley & Sons.

Badalamenti, Jonathan P., Zarath M. Summers, Chi Ho Chan, Jeffrey A. Gralnick, and Daniel R. Bond. 2016. "Isolation and Genomic Characterization of 'Desulfuromonas Soudanensis WTL', a Metal- and Electrode-Respiring Bacterium from Anoxic Deep Subsurface Brine." *Frontiers in Microbiology* 7. doi: 10.3389/fmicb.2016.00913.

Bahulikar, Rahul A., Ivone Torres-Jerez, Eric Worley, Kelly Craven, and Michael K. Udvardi. 2014. "Diversity of Nitrogen-Fixing Bacteria Associated with Switchgrass in the Native Tallgrass Prairie of Northern Oklahoma." *Applied and Environmental Microbiology* 80(18):5636–43. doi: 10.1128/AEM.02091-14.

Bailey, Alexis C., Matthew Kellom, Amisha T. Poret-Peterson, Kathryn Noonan, Hilairy E. Hartnett, and Jason Raymond. 2014. "Draft Genome Sequence of Massilia Sp. Strain BSC265, Isolated from Biological Soil Crust of Moab, Utah." *Genome Announcements* 2(6):e01199-14. doi: 10.1128/genomeA.01199-14.

Banerjee, Meenakshi, and Vidhi Verma. 2009. "Nitrogen Fixation in Endolithic Cyanobacterial Communities of the McMurdo Dry Valley, Antarctica." *ScienceAsia* 35(3):215. doi: 10.2306/scienceasia1513-1874.2009.35.215.

Beijerinck, Martinus W. 1888. "Die Bacterien Der Papilionaceenknöllchen." *Botanische Zeitung* 46(46):725–35.

Bell, Terrence H., Etienne Yergeau, Christine Martineau, David Juck, Lyle G. Whyte, and Charles W. Greer. 2011. “Identification of Nitrogen-Incorporating Bacteria in Petroleum-Contaminated Arctic Soils by Using [15N]DNA-Based Stable Isotope Probing and Pyrosequencing.” *Applied and Environmental Microbiology* 77(12):4163–71. doi: 10.1128/AEM.00172-11.

Belnap, Jayne. 2002. “Nitrogen Fixation in Biological Soil Crusts from Southeast Utah, USA.” *Biology and Fertility of Soils* 35(2):128–35. doi: 10.1007/s00374-002-0452-x.

Bentley, S. D., K. F. Chater, A. M. Cerdeño-Tárraga, G. L. Challis, N. R. Thomson, K. D. James, D. E. Harris, M. A. Quail, H. Kieser, D. Harper, A. Bateman, S. Brown, G. Chandra, C. W. Chen, M. Collins, A. Cronin, A. Fraser, A. Goble, J. Hidalgo, T. Hornsby, S. Howarth, C. H. Huang, T. Kieser, L. Larke, L. Murphy, K. Oliver, S. O’Neil, E. Rabbinowitsch, M. A. Rajandream, K. Rutherford, S. Rutter, K. Seeger, D. Saunders, S. Sharp, R. Squares, S. Squares, K. Taylor, T. Warren, A. Wietzorrek, J. Woodward, B. G. Barrell, J. Parkhill, and D. A. Hopwood. 2002. “Complete Genome Sequence of the Model Actinomycete Streptomyces Coelicolor A3(2).” *Nature* 417(6885):141–47. doi: 10.1038/417141a.

Bérdy, János. 2012. “Thoughts and Facts about Antibiotics: Where We Are Now and Where We Are Heading.” *The Journal of Antibiotics* 65(8):385–95. doi: 10.1038/ja.2012.27.

Berg, Gabriele, Nicolle Roskot, Anette Steidle, Leo Eberl, Angela Zock, and Kornelia Smalla. 2002. “Plant-Dependent Genotypic and Phenotypic Diversity of Antagonistic Rhizobacteria Isolated from Different Verticillium Host Plants.” *Applied and Environmental Microbiology* 68(7):3328–38. doi: 10.1128/AEM.68.7.3328-3338.2002.

Bertsova, Y. V., A. V. Bogachev, and V. P. Skulachev. 2001. “Noncoupled NADH:Ubiquinone Oxidoreductase of Azotobacter Vinelandii Is Required for Diazotrophic Growth at High Oxygen Concentrations.” *Journal of Bacteriology* 183(23):6869–74. doi: 10.1128/JB.183.23.6869-6874.2001.

Bhaya, Devaki, Arthur R. Grossman, Anne-Soisig Steunou, Natalia Khuri, Frederick M. Cohan, Natsuko Hamamura, Melanie C. Melendrez, Mary M. Bateson, David M. Ward, and John F. Heidelberg. 2007. “Population Level Functional Diversity in a Microbial Community Revealed by Comparative Genomic and Metagenomic Analyses.” *The ISME Journal* 1(8):703–13. doi: 10.1038/ismej.2007.46.

Biabani, Abbas. 2008. “Endosymbiosis Formation Between a Nitrogen-Fixing Mix of Bacteria (Xanthomonas Sp. + Arthrobacter Sp.) and Wheat Root Cells.” doi: 10.1007/978-1-4020-8252-8_113.

Blasco, Rafael, Manuel Martínez-Luque, Marta P. Madrid, Francisco Castillo, and Conrado Moreno-Vivián. 2001. "Rhodococcus Sp. RB1 Grows in the Presence of High Nitrate and Nitrite Concentrations and Assimilates Nitrate in Moderately Saline Environments." *Archives of Microbiology* 175(6):435–40. doi: 10.1007/s002030100285.

Borodina, Irina, Preben Krabben, and Jens Nielsen. 2005. "Genome-Scale Analysis of Streptomyces Coelicolor A3(2) Metabolism." *Genome Research* 15(6):820–29. doi: 10.1101/gr.3364705.

Bothe, Hermann, Stuart Ferguson, and William E. Newton. 2006. *Biology of the Nitrogen Cycle*. Elsevier.

Boyd, Eric, and John W. Peters. 2013. "New Insights into the Evolutionary History of Biological Nitrogen Fixation." *Frontiers in Microbiology* 4. doi: 10.3389/fmicb.2013.00201.

Bruns, Thomas D., Robert Fogel, and John W. Taylor. 1990. "Amplification and Sequencing of Dna from Fungal Herbarium Specimens." *Mycologia* 82(2):175–84. doi: 10.1080/00275514.1990.12025863.

Buck, Martin, Stephen Miller, Martin Drummond, and Ray Dixon. 1986. "Upstream Activator Sequences Are Present in the Promoters of Nitrogen Fixation Genes." *Nature* 320(6060):374–78. doi: 10.1038/320374a0.

Bulen, W. A., and J. R. LeComte. 1966. "The Nitrogenase System from Azotobacter: Two-Enzyme Requirement for N2 Reduction, ATP-Dependent H2 Evolution, and ATP Hydrolysis." *Proceedings of the National Academy of Sciences of the United States of America* 56(3):979–86. doi: 10.1073/pnas.56.3.979.

Burén, Stefan, Emilio Jiménez-Vicente, Carlos Echavarri-Erasun, and Luis M. Rubio. 2020. "Biosynthesis of Nitrogenase Cofactors." *Chemical Reviews* 120(12):4921–68. doi: 10.1021/acs.chemrev.9b00489.

Burgess, Barbara K., and David J. Lowe. 1996. "Mechanism of Molybdenum Nitrogenase." *Chemical Reviews* 96(7):2983–3012. doi: 10.1021/cr950055x.

Cabello, Purificación, M. Dolores Roldán, and Conrado Moreno-Vivián. 2004. "Nitrate Reduction and the Nitrogen Cycle in Archaea." *Microbiology (Reading, England)* 150(Pt 11):3527–46. doi: 10.1099/mic.0.27303-0.

Çakmakçı, Ramazan, Mustafa Erat, Ümmügülsüm Erdoğan, and Mesude Figen Dönmez. 2007a. "The Influence of Plant Growth–Promoting Rhizobacteria on Growth and Enzyme Activities in Wheat and Spinach Plants." *Journal of Plant Nutrition and Soil Science* 170(2):288–95. doi: https://doi.org/10.1002/jpln.200625105.

Çakmakçı, Ramazan, Mustafa Erat, Ümmügülsüm Erdoğan, and Mesude Figen Dönmez. 2007b. "The Influence of Plant Growth–Promoting Rhizobacteria on Growth and Enzyme Activities in Wheat and Spinach Plants." *Journal of Plant Nutrition and Soil Science* 170(2):288–95. doi: https://doi.org/10.1002/jpln.200625105.

Cardenas, Erick, and James M. Tiedje. 2008. "New Tools for Discovering and Characterizing Microbial Diversity." *Current Opinion in Biotechnology* 19(6):544–49. doi: 10.1016/j.copbio.2008.10.010.

Challis, Gregory L., and David A. Hopwood. 2003. "Synergy and Contingency as Driving Forces for the Evolution of Multiple Secondary Metabolite Production by Streptomyces Species." *Proceedings of the National Academy of Sciences of the United States of America* 100 Suppl 2:14555–61. doi: 10.1073/pnas.1934677100.

Che, Rongxiao, Yongcui Deng, Fang Wang, Weijin Wang, Zhihong Xu, Yanbin Hao, Kai Xue, Biao Zhang, Li Tang, Huakun Zhou, and Xiaoyong Cui. 2018. "Autotrophic and Symbiotic Diazotrophs Dominate Nitrogen-Fixing Communities in Tibetan Grassland Soils." *Science of The Total Environment* 639:997–1006. doi: 10.1016/j.scitotenv.2018.05.238.

Chen, Lin-xing, Min Hu, Li-nan Huang, Zheng-shuang Hua, Jia-liang Kuang, Sheng-jin Li, and Wen-sheng Shu. 2015. "Comparative Metagenomic and Metatranscriptomic Analyses of Microbial Communities in Acid Mine Drainage." *The ISME Journal* 9(7):1579–92. doi: 10.1038/ismej.2014.245.

Chien, Yueh-Tyng, and Stephen H. Zinder. 1996. "Cloning, Functional Organization, Transcript Studies, and Phylogenetic Analysis of the Complete Nitrogenase Structural Genes (NifHDK2) and Associated Genes in the Archaeon Methanosarcina Barkeri 227." *Journal of Bacteriology* 178(1):143–48.

Cloutier, Mara Lee, Arnab Bhowmik, Terrence H. Bell, and Mary Ann Bruns. 2019. "Innovative Technologies Can Improve Understanding of Microbial Nitrogen Dynamics in Agricultural Soils." *Agricultural & Environmental Letters* 4(1):190032. doi: 10.2134/ael2019.08.0032.

Cooper, Geoffrey M. 2000. "The Molecular Composition of Cells." *The Cell: A Molecular Approach. 2nd Edition*.

Dahal, Bibha, Gitanjali NandaKafle, Lora Perkins, and Volker S. Brözel. 2017. "Diversity of Free-Living Nitrogen Fixing Streptomyces in Soils of the Badlands of South Dakota." *Microbiological Research* 195:31–39. doi: 10.1016/j.micres.2016.11.004.

Danczak, R. E., M. D. Johnston, C. Kenah, M. Slattery, K. C. Wrighton, and M. J. Wilkins. 2017. "Members of the Candidate Phyla Radiation Are Functionally Differentiated by Carbon- and Nitrogen-Cycling Capabilities." *Microbiome* 5. doi: 10.1186/s40168-017-0331-1.

Davis, Sarah C., William J. Parton, Frank G. Dohleman, Candice M. Smith, Stephen Del Grosso, Angela D. Kent, and Evan H. DeLucia. 2010. "Comparative Biogeochemical Cycles of Bioenergy Crops Reveal Nitrogen-Fixation and Low Greenhouse Gas Emissions in a Miscanthus × Giganteus Agro-Ecosystem." *Ecosystems* 13(1):144–56. doi: 10.1007/s10021-009-9306-9.

Delmont, Tom O., Christopher Quince, Alon Shaiber, Özcan C. Esen, Sonny TM Lee, Michael S. Rappé, Sandra L. McLellan, Sebastian Lücker, and A. Murat Eren. 2018. "Nitrogen-Fixing Populations of Planctomycetes and Proteobacteria Are Abundant in Surface Ocean Metagenomes." *Nature Microbiology* 3(7):804–13. doi: 10.1038/s41564-018-0176-9.

Desai, Mahesh S., Karoline Assig, and Sharmishtha Dattagupta. 2013. "Nitrogen Fixation in Distinct Microbial Niches within a Chemoautotrophy-Driven Cave Ecosystem." *The ISME Journal* 7(12):2411–23. doi: 10.1038/ismej.2013.126.

Ding J., Sun H., Su F., Xu Q., and Huang Y. 1981. "Studies on nitrogen fixation by actinomycetes [China]." *Wei sheng wu hsueh pao = Acta microbiologica sinica*.

Ding, Y., J. Wang, Y. Liu, and S. Chen. 2005. "Isolation and Identification of Nitrogen-Fixing Bacilli from Plant Rhizospheres in Beijing Region." *Journal of Applied Microbiology* 99(5):1271–81. doi: 10.1111/j.1365-2672.2005.02738.x.

Dingler, C., J. Kuhla, H. Wassink, and J. Oelze. 1988. "Levels and Activities of Nitrogenase Proteins in Azotobacter Vinelandii Grown at Different Dissolved Oxygen Concentrations." *Journal of Bacteriology* 170(5):2148–52.

Dixon, Ray, and Daniel Kahn. 2004a. "Genetic Regulation of Biological Nitrogen Fixation." *Nature Reviews Microbiology* 2(8):621–31. doi: 10.1038/nrmicro954.

Dixon, Ray, and Daniel Kahn. 2004b. "Genetic Regulation of Biological Nitrogen Fixation." *Nature Reviews Microbiology* 2(8):621–31. doi: 10.1038/nrmicro954.

Dobereiner, J., I. E. Marriel, and M. Nery. 1976. "Ecological Distribution of Spirillum Lipoferum Beijerinck." *Canadian Journal of Microbiology* 22(10):1464–73. doi: 10.1139/m76-217.

Dos Santos, Patricia C., Zhong Fang, Steven W. Mason, João C. Setubal, and Ray Dixon. 2012. "Distribution of Nitrogen Fixation and Nitrogenase-like Sequences amongst Microbial Genomes." *BMC Genomics* 13(1):162. doi: 10.1186/1471-2164-13-162.

Doty, Sharon L., Brian Oakley, Gang Xin, Jun Won Kang, Glenda Singleton, Zareen Khan, Azra Vajzovic, and James T. Staley. 2009. "Diazotrophic Endophytes of Native Black Cottonwood and Willow." *Symbiosis* 47(1):23–33. doi: 10.1007/BF03179967.

Eady, Robert R. 1996. “Structure−Function Relationships of Alternative Nitrogenases.” *Chemical Reviews* 96(7):3013–30. doi: 10.1021/cr950057h.

Eady, Robert R. 2003. “Current Status of Structure Function Relationships of Vanadium Nitrogenase.” *Coordination Chemistry Reviews* 237(1):23–30. doi: 10.1016/S0010-8545(02)00248-5.

Eelen, Guy, Charlotte Dubois, Anna Rita Cantelmo, Jermaine Goveia, Ulrike Brüning, Michael DeRan, Gopala Jarugumilli, Jos van Rijssel, Giorgio Saladino, Federico Comitani, Annalisa Zecchin, Susana Rocha, Rongyuan Chen, Hongling Huang, Saar Vandekeere, Joanna Kalucka, Christian Lange, Francisco Morales-Rodriguez, Bert Cruys, Lucas Treps, Leanne Ramer, Stefan Vinckier, Katleen Brepoels, Sabine Wyns, Joris Souffreau, Luc Schoonjans, Wouter H. Lamers, Yi Wu, Jurgen Haustraete, Johan Hofkens, Sandra Liekens, Richard Cubbon, Bart Ghesquière, Mieke Dewerchin, Francesco L. Gervasio, Xuri Li, Jaap D. van Buul, Xu Wu, and Peter Carmeliet. 2018. “Role of Glutamine Synthetase in Angiogenesis beyond Glutamine Synthesis.” *Nature* 561(7721):63–69. doi: 10.1038/s41586-018-0466-7.

Esteves-Ferreira, Alberto A., João Henrique Frota Cavalcanti, Marcelo Gomes Marçal Vieira Vaz, Luna V. Alvarenga, Adriano Nunes-Nesi, and Wagner L. Araújo. 2017. “Cyanobacterial Nitrogenases: Phylogenetic Diversity, Regulation and Functional Predictions.” *Genetics and Molecular Biology* 40(1 suppl 1):261–75. doi: 10.1590/1678-4685-gmb-2016-0050.

Estrada-de Los Santos, Paulina, Marike Palmer, Belén Chávez-Ramírez, Chrizelle Beukes, Emma T. Steenkamp, Leah Briscoe, Noor Khan, Marta Maluk, Marcel Lafos, Ethan Humm, Monique Arrabit, Matthew Crook, Eduardo Gross, Marcelo F. Simon, Fábio Bueno Dos Reis Junior, William B. Whitman, Nicole Shapiro, Philip S. Poole, Ann M. Hirsch, Stephanus N. Venter, and Euan K. James. 2018. “Whole Genome Analyses Suggests That Burkholderia Sensu Lato Contains Two Additional Novel Genera (Mycetohabitans Gen. Nov., and Trinickia Gen. Nov.): Implications for the Evolution of Diazotrophy and Nodulation in the Burkholderiaceae.” *Genes* 9(8). doi: 10.3390/genes9080389.

Fang, Kai, Zhu-Shou-Neng Bao, Lin Chen, Jie Zhou, Zhi-Ping Yang, Xing-Fan Dong, and Han-Bo Zhang. 2019. “Growth-Promoting Characteristics of Potential Nitrogen-Fixing Bacteria in the Root of an Invasive Plant Ageratina Adenophora.” *PeerJ* 7:e7099. doi: 10.7717/peerj.7099.

Felsenstein, Joseph. 1993. *PHYLIP (Phylogeny Inference Package), Version 3.5 c*. Joseph Felsenstein.

Feng, Haiwei, Yujing Sun, Yuee Zhi, Xing Wei, Yanqing Luo, Liang Mao, and Pei Zhou. 2014. “Identification and Characterization of the Nitrate Assimilation Genes in

the Isolate of Streptomyces Griseorubens JSD-1." *Microbial Cell Factories* 13(1):174. doi: 10.1186/s12934-014-0174-4.

Feng, Maomao, Jonathan M. Adams, Kunkun Fan, Yu Shi, Ruibo Sun, Daozhong Wang, Xisheng Guo, and Haiyan Chu. 2018. "Long-Term Fertilization Influences Community Assembly Processes of Soil Diazotrophs." *Soil Biology and Biochemistry* 126:151–58. doi: 10.1016/j.soilbio.2018.08.021.

Fischer, Marco, and R. Gary Sawers. 2013. "A Universally Applicable and Rapid Method for Measuring the Growth of Streptomyces and Other Filamentous Microorganisms by Methylene Blue Adsorption-Desorption." *Applied and Environmental Microbiology* 79(14):4499–4502. doi: 10.1128/AEM.00778-13.

Forchhammer, Karl. 2007. "Glutamine Signaling in Bacteria." *Frontiers in Bioscience : A Journal and Virtual Library* 12:358–70. doi: 10.2741/2069.

Fowler, D., C. E. Steadman, D. Stevenson, M. Coyle, R. M. Rees, U. M. Skiba, M. A. Sutton, J. N. Cape, A. J. Dore, M. Vieno, D. Simpson, S. Zaehle, B. D. Stocker, M. Rinaldi, M. C. Facchini, C. R. Flechard, E. Nemitz, M. Twigg, J. W. Erisman, K. Butterbach-Bahl, and J. N. Galloway. 2015. "Effects of Global Change during the 21st Century on the Nitrogen Cycle." *Atmospheric Chemistry and Physics* 15(24):13849–93. doi: 10.5194/acp-15-13849-2015.

Fowler, David, Mhairi Coyle, Ute Skiba, Mark A. Sutton, J. Neil Cape, Stefan Reis, Lucy J. Sheppard, Alan Jenkins, Bruna Grizzetti, James N. Galloway, Peter Vitousek, Allison Leach, Alexander F. Bouwman, Klaus Butterbach-Bahl, Frank Dentener, David Stevenson, Marcus Amann, and Maren Voss. 2013. "The Global Nitrogen Cycle in the Twenty-First Century." *Philosophical Transactions of the Royal Society B: Biological Sciences* 368(1621):20130164. doi: 10.1098/rstb.2013.0164.

Fuentes-Ramírez, Luis E., Rocio Bustillos-Cristales, Armando Tapia-Hernández, Teresita Jiménez-Salgado, E. Tao Wang, Esperanza Martínez-Romero, and Jesus Caballero-Mellado. 2001. "Novel Nitrogen-Fixing Acetic Acid Bacteria, Gluconacetobacter Johannae Sp. Nov. and Gluconacetobacter Azotocaptans Sp. Nov., Associated with Coffee Plants." *International Journal of Systematic and Evolutionary Microbiology* 51(4):1305–14.

Fujita, Yuichi, and Carl 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 *." *Journal of Biological Chemistry* 275(31):23583–88. doi: 10.1074/jbc.M002904200.

Gaby, John Christian, and Daniel H. Buckley. 2011. "A Global Census of Nitrogenase Diversity." *Environmental Microbiology* 13(7):1790–99. doi: 10.1111/j.1462-2920.2011.02488.x.

Gaby, John Christian, and Daniel H. Buckley. 2012a. "A Comprehensive Evaluation of PCR Primers to Amplify the NifH Gene of Nitrogenase." *PLOS ONE* 7(7):e42149. doi: 10.1371/journal.pone.0042149.

Gaby, John Christian, and Daniel H. Buckley. 2012b. "A Comprehensive Evaluation of PCR Primers to Amplify the NifH Gene of Nitrogenase." *PloS One* 7(7):e42149. doi: 10.1371/journal.pone.0042149.

Gadkari, D., G. Mörsdorf, and O. Meyer. 1992. "Chemolithoautotrophic Assimilation of Dinitrogen by Streptomyces Thermoautotrophicus UBT1: Identification of an Unusual N2-Fixing System." *Journal of Bacteriology* 174(21):6840–43. doi: 10.1128/jb.174.21.6840-6843.1992.

Gallon, J. R. 1981. "The Oxygen Sensitivity of Nitrogenase: A Problem for Biochemists and Micro-Organisms." *Trends in Biochemical Sciences* 6:19–23. doi: 10.1016/0968-0004(81)90008-6.

Geiss, Gary K., Roger E. Bumgarner, Brian Birditt, Timothy Dahl, Naeem Dowidar, Dwayne L. Dunaway, H. Perry Fell, Sean Ferree, Renee D. George, Tammy Grogan, Jeffrey J. James, Malini Maysuria, Jeffrey D. Mitton, Paola Oliveri, Jennifer L. Osborn, Tao Peng, Amber L. Ratcliffe, Philippa J. Webster, Eric H. Davidson, Leroy Hood, and Krassen Dimitrov. 2008. "Direct Multiplexed Measurement of Gene Expression with Color-Coded Probe Pairs." *Nature Biotechnology* 26(3):317–25. doi: 10.1038/nbt1385.

González, P. J., C. Correia, Isabel Moura, C. D. Brondino, and J. J. G. Moura. 2006. "Bacterial Nitrate Reductases: Molecular and Biological Aspects of Nitrate Reduction." *Journal of Inorganic Biochemistry* 100(5–6):1015–23. doi: 10.1016/j.jinorgbio.2005.11.024.

Gopalakrishnan, S., V. Srinivas, and S. L. Prasanna. 2020. "Streptomyces." Pp. 55–71 in *Beneficial Microbes in Agro-Ecology*. Elsevier.

Graber, Joseph R., Jared R. Leadbetter, and John A. Breznak. 2004. "Description of Treponema Azotonutricium Sp. Nov. and Treponema Primitia Sp. Nov., the First Spirochetes Isolated from Termite Guts." *Applied and Environmental Microbiology* 70(3):1315–20. doi: 10.1128/AEM.70.3.1315-1320.2004.

Grabherr, Manfred G., Brian J. Haas, Moran Yassour, Joshua Z. Levin, Dawn A. Thompson, Ido Amit, Xian Adiconis, Lin Fan, Raktima Raychowdhury, Qiandong Zeng, Zehua Chen, Evan Mauceli, Nir Hacohen, Andreas Gnirke, Nicholas Rhind, Federica di Palma, Bruce W. Birren, Chad Nusbaum, Kerstin Lindblad-Toh, Nir Friedman, and Aviv Regev. 2011. "Trinity: Reconstructing a Full-Length Transcriptome without a Genome from RNA-Seq Data." *Nature Biotechnology* 29(7):644–52. doi: 10.1038/nbt.1883.

Gtari, Maher, Faten Ghodhbane-Gtari, Imen Nouioui, Nicholas Beauchemin, and Louis S. Tisa. 2012. “Phylogenetic Perspectives of Nitrogen-Fixing Actinobacteria.” *Archives of Microbiology* 194(1):3–11. doi: 10.1007/s00203-011-0733-6.

Guindon, Stéphane, Jean-François Dufayard, Vincent Lefort, Maria Anisimova, Wim Hordijk, and Olivier Gascuel. 2010. “New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0.” *Systematic Biology* 59(3):307–21. doi: 10.1093/sysbio/syq010.

Guo, He-Bao, Shan-Wen He, Xing Wang, Kyu-Kyu Thin, Hai-Lei Wei, and Xiao-Xia Zhang. 2020. “Bacillus Telluris Sp. Nov. Isolated from Greenhouse Soil in Beijing, China.” *Microorganisms* 8(5):E702. doi: 10.3390/microorganisms8050702.

Halsall, Dorothy M., and Alan H. Gibson. 1985. “Cellulose Decomposition and Associated Nitrogen Fixation by Mixed Cultures of Cellulomonas Gelida and Azospirillum Species or Bacillus Macerans.” *Applied and Environmental Microbiology* 50(4):1021–26.

Hartemink, Alfred E., and J. G. Bockheim. 2013. “Soil Genesis and Classification.” *CATENA* 104:251–56. doi: 10.1016/j.catena.2012.12.001.

Hatayama, Kouta, Satomi Kawai, Hirofumi Shoun, Yasuichi Ueda, and Akira Nakamura. 2005. “Pseudomonas Azotifigens Sp. Nov., a Novel Nitrogen-Fixing Bacterium Isolated from a Compost Pile.” *International Journal of Systematic and Evolutionary Microbiology* 55(Pt 4):1539–44. doi: 10.1099/ijs.0.63586-0.

Hayatsu, Masahito, Kanako Tago, and Masanori Saito. 2008. “Various Players in the Nitrogen Cycle: Diversity and Functions of the Microorganisms Involved in Nitrification and Denitrification.” *Soil Science and Plant Nutrition* 54(1):33–45. doi: 10.1111/j.1747-0765.2007.00195.x.

Hayden, Helen L., Judy Drake, Mark Imhof, Andrew P. A. Oxley, Sorn Norng, and Pauline M. Mele. 2010. “The Abundance of Nitrogen Cycle Genes AmoA and NifH Depends on Land-Uses and Soil Types in South-Eastern Australia.” *Soil Biology and Biochemistry* 42(10):1774–83. doi: 10.1016/j.soilbio.2010.06.015.

Hellriegel, Hermann, and Hermann Wilfarth. 1888. *Untersuchungen Über Die Stickstoffnahrung Der Gramineen Und Leguminosen*. Berlin,: Buchdruckerei der “Post” Kayssler,.

Héry, Marina, Laurent Philippot, Eve Mériaux, Franck Poly, Xavier Le Roux, and Elisabeth Navarro. 2005. “Nickel Mine Spoils Revegetation Attempts: Effect of Pioneer Plants on Two Functional Bacterial Communities Involved in the N-Cycle.” *Environmental Microbiology* 7(4):486–98. doi: 10.1111/j.1462-2920.2005.00705.x.

Higdon, Shawn M., Bihua C. Huang, Alan B. Bennett, and Bart C. Weimer. 2020. "Identification of Nitrogen Fixation Genes in Lactococcus Isolated from Maize Using Population Genomics and Machine Learning." *Microorganisms* 8(12). doi: 10.3390/microorganisms8122043.

Higdon, Shawn M., Tania Pozzo, Nguyet Kong, Bihua C. Huang, Mai Lee Yang, Richard Jeannotte, C. Titus Brown, Alan B. Bennett, and Bart C. Weimer. 2020. "Genomic Characterization of a Diazotrophic Microbiota Associated with Maize Aerial Root Mucilage" edited by J.-T. Chen. *PLOS ONE* 15(9):e0239677. doi: 10.1371/journal.pone.0239677.

Hinnemann, Berit, and Jens K. Nørskov. 2004. "Structure of the FeFe-Cofactor of the Iron-Only Nitrogenase and Possible Mechanism for Dinitrogen Reduction." *Physical Chemistry Chemical Physics* 6(4):843–53. doi: 10.1039/B310850C.

Hoffman, Brian M., Dmitriy Lukoyanov, Zhi-Yong Yang, Dennis R. Dean, and Lance C. Seefeldt. 2014. "Mechanism of Nitrogen Fixation by Nitrogenase: The Next Stage." *Chemical Reviews* 114(8):4041–62. doi: 10.1021/cr400641x.

Houlton, B. Z., S. L. Morford, and R. A. Dahlgren. 2018. "Convergent Evidence for Widespread Rock Nitrogen Sources in Earth's Surface Environment." *Science* 360(6384):58–62. doi: 10.1126/science.aan4399.

Howard, James B., and Douglas C. Rees. 1996. "Structural Basis of Biological Nitrogen Fixation." *Chemical Reviews* 96(7):2965–82. doi: 10.1021/cr9500545.

Howard, James B., and Douglas C. Rees. 2006. "How Many Metals Does It Take to Fix N2? A Mechanistic Overview of Biological Nitrogen Fixation." *Proceedings of the National Academy of Sciences* 103(46):17088–93. doi: 10.1073/pnas.0603978103.

Hu, Yilin, Chi Chung Lee, and Markus W. Ribbe. 2012. "Vanadium Nitrogenase: A Two-Hit Wonder?" *Dalton Transactions* 41(4):1118–27. doi: 10.1039/C1DT11535A.

Hu, Yilin, and Markus W. Ribbe. 2014. "A Journey into the Active Center of Nitrogenase." *Journal of Biological Inorganic Chemistry : JBIC : A Publication of the Society of Biological Inorganic Chemistry* 19(6):731–36. doi: 10.1007/s00775-014-1137-2.

Hu, Yilin, and Markus W. Ribbe. 2015. "Nitrogenase and Homologs." *Journal of Biological Inorganic Chemistry: JBIC: A Publication of the Society of Biological Inorganic Chemistry* 20(2):435–45. doi: 10.1007/s00775-014-1225-3.

Hug, Laura A., Brett J. Baker, Karthik Anantharaman, Christopher T. Brown, Alexander J. Probst, Cindy J. Castelle, Cristina N. Butterfield, Alex W. Hernsdorf, Yuki Amano, Kotaro Ise, Yohey Suzuki, Natasha Dudek, David A. Relman, Kari M.

Finstad, Ronald Amundson, Brian C. Thomas, and Jillian F. Banfield. 2016. "A New View of the Tree of Life." *Nature Microbiology* 1(5):1–6. doi: 10.1038/nmicrobiol.2016.48.

Ikbal, Nishat Passricha, Shabnam K. Saifi, Virendra K. Sikka, and Narendra Tuteja. 2020. "Chapter 3 - Multilegume Biofertilizer: A Dream." Pp. 35–45 in *Molecular Aspects of Plant Beneficial Microbes in Agriculture*, edited by V. Sharma, R. Salwan, and L. K. T. Al-Ani. Academic Press.

Inoue, Jun-ichi, Kenshiro Oshima, Wataru Suda, Mitsuo Sakamoto, Takao Iino, Satoko Noda, Yuichi Hongoh, Masahira Hattori, and Moriya Ohkuma. 2015. "Distribution and Evolution of Nitrogen Fixation Genes in the Phylum Bacteroidetes." *Microbes and Environments* 30(1):44–50. doi: 10.1264/jsme2.ME14142.

Isobe, Kazuo, Keisuke Koba, Shigeto Otsuka, and Keishi Senoo. 2011. "Nitrification and Nitrifying Microbial Communities in Forest Soils." *Journal of Forest Research* 16(5):351. doi: 10.1007/s10310-011-0266-5.

Isobe, Kazuo, and Nobuhito Ohte. 2014. "Ecological Perspectives on Microbes Involved in N-Cycling." *Microbes and Environments* 29(1):4–16. doi: 10.1264/jsme2.ME13159.

Issa, Ahmed A., Mohamed Hemida Abd-Alla, and Takuji Ohyama. 2014. "Nitrogen Fixing Cyanobacteria: Future Prospect." *Advances in Biology and Ecology of Nitrogen Fixation*. doi: 10.5772/56995.

Iwata, Kenichi, Azlin Azlan, Hironori Yamakawa, and Toshio Omori. 2010. "Ammonia Accumulation in Culture Broth by the Novel Nitrogen-Fixing Bacterium, Lysobacter Sp. E4." *Journal of Bioscience and Bioengineering* 110(4):415–18. doi: 10.1016/j.jbiosc.2010.05.006.

Jacobson, M. R., K. E. Brigle, L. T. Bennett, R. A. Setterquist, M. S. Wilson, V. L. Cash, J. Beynon, W. E. Newton, and D. R. Dean. 1989. "Physical and Genetic Map of the Major Nif Gene Cluster from Azotobacter Vinelandii." *Journal of Bacteriology* 171(2):1017–27. doi: 10.1128/jb.171.2.1017-1027.1989.

Jang, S. B., L. C. Seefeldt, and J. W. Peters. 2000. "Insights into Nucleotide Signal Transduction in Nitrogenase: Structure of an Iron Protein with MgADP Bound." *Biochemistry* 39(48):14745–52. doi: 10.1021/bi001705g.

Jewell, Talia N. M., Ulas Karaoz, Eoin L. Brodie, Kenneth H. Williams, and Harry R. Beller. 2016. "Metatranscriptomic Evidence of Pervasive and Diverse Chemolithoautotrophy Relevant to C, S, N and Fe Cycling in a Shallow Alluvial Aquifer." *The ISME Journal* 10(9):2106–17. doi: 10.1038/ismej.2016.25.

Joshi, Divya, Ramesh Chandra, Deep Chandra Suyal, Saurabh Kumar, and Reeta Goel. 2019a. “Impacts of Bioinoculants Pseudomonas Jesenii MP1 and Rhodococcus Qingshengii S10107 on Chickpea (Cicer Arietinum L.) Yield and Soil Nitrogen Status.” *Pedosphere* 29(3):388–99. doi: 10.1016/S1002-0160(19)60807-6.

Joshi, Divya, Ramesh Chandra, Deep Chandra Suyal, Saurabh Kumar, and Reeta Goel. 2019b. “Impacts of Bioinoculants Pseudomonas Jesenii MP1 and Rhodococcus Qingshengii S10107 on Chickpea (Cicer Arietinum L.) Yield and Soil Nitrogen Status.” *Pedosphere* 29(3):388–99. doi: 10.1016/S1002-0160(19)60807-6.

Kämpfer, Peter, Hans-Jürgen Busse, John A. McInroy, Jia Xu, and Stefanie P. Glaeser. 2015. “Flavobacterium Nitrogenifigens Sp. Nov., Isolated from Switchgrass (Panicum Virgatum).” *International Journal of Systematic and Evolutionary Microbiology* 65(9):2803–9. doi: 10.1099/ijs.0.000330.

Kanehisa, M., and S. Goto. 2000. “KEGG: Kyoto Encyclopedia of Genes and Genomes.” *Nucleic Acids Research* 28(1):27–30. doi: 10.1093/nar/28.1.27.

Kanehisa, Minoru, and Yoko Sato. 2020. “KEGG Mapper for Inferring Cellular Functions from Protein Sequences.” *Protein Science: A Publication of the Protein Society* 29(1):28–35. doi: 10.1002/pro.3711.

Kanehisa, Minoru, Yoko Sato, and Kanae Morishima. 2016. “BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences.” *Journal of Molecular Biology* 428(4):726–31. doi: 10.1016/j.jmb.2015.11.006.

Kapili, Bennett J., Samuel E. Barnett, Daniel H. Buckley, and Anne E. Dekas. 2020. “Evidence for Phylogenetically and Catabolically Diverse Active Diazotrophs in Deep-Sea Sediment.” *The ISME Journal* 14(4):971–83. doi: 10.1038/s41396-019-0584-8.

Kasting, J. F., and James C. G. Walker. 1981. “Limits on Oxygen Concentration in the Prebiological Atmosphere and the Rate of Abiotic Fixation of Nitrogen.” *Journal of Geophysical Research: Oceans* 86(C2):1147–58. doi: https://doi.org/10.1029/JC086iC02p01147.

van Kessel, Maartje A. H. J., Daan R. Speth, Mads Albertsen, Per H. Nielsen, Huub J. M. Op den Camp, Boran Kartal, Mike S. M. Jetten, and Sebastian Lücker. 2015. “Complete Nitrification by a Single Microorganism.” *Nature* 528(7583):555–59. doi: 10.1038/nature16459.

Khadem, Ahmad F., Arjan Pol, Mike S. M. Jetten, and Huub J. M. Op den Camp. 2010. “Nitrogen Fixation by the Verrucomicrobial Methanotroph ‘Methylacidiphilum Fumariolicum’ SolV.” *Microbiology (Reading, England)* 156(Pt 4):1052–59. doi: 10.1099/mic.0.036061-0.

Kim, Daehwan, Geo Pertea, Cole Trapnell, Harold Pimentel, Ryan Kelley, and Steven L. Salzberg. 2013. “TopHat2: Accurate Alignment of Transcriptomes in the Presence of Insertions, Deletions and Gene Fusions.” *Genome Biology* 14(4):R36. doi: 10.1186/gb-2013-14-4-r36.

Kim, Jongsun, and Douglas C. Rees. 1994. “Nitrogenase and Biological Nitrogen Fixation.” *Biochemistry* 33(2):389–97. doi: 10.1021/bi00168a001.

Klipp, Werner, Bernd Masepohl, John R. Gallon, and William E. Newton, eds. 2005. *Genetics and Regulation of Nitrogen Fixation in Free-Living Bacteria*. Vol. 2. Dordrecht: Kluwer Academic Publishers.

Konieczna, Iwona, Paulina Żarnowiec, Marek Kwinkowski, Beata Kolesińska, Justyna Frączyk, Zbigniew Kamiński, and Wiesław Kaca. 2012. “Bacterial Urease and Its Role in Long-Lasting Human Diseases.” *Current Protein & Peptide Science* 13(8):789–806. doi: 10.2174/138920312804871094.

König, Sten, Olivier Gros, Stefan E. Heiden, Tjorven Hinzke, Andrea Thürmer, Anja Poehlein, Susann Meyer, Magalie Vatin, Didier Mbéguié-A-Mbéguié, Jennifer Tocny, Ruby Ponnudurai, Rolf Daniel, Dörte Becher, Thomas Schweder, and Stephanie Markert. 2016. “Nitrogen Fixation in a Chemoautotrophic Lucinid Symbiosis.” *Nature Microbiology* 2(1):1–10. doi: 10.1038/nmicrobiol.2016.193.

Kruse, Thomas, Chandini Murarilal Ratnadevi, Helge-André Erikstad, and Nils-Kåre Birkeland. 2019. “Complete Genome Sequence Analysis of the Thermoacidophilic Verrucomicrobial Methanotroph ‘Candidatus Methylacidiphilum Kamchatkense’ Strain Kam1 and Comparison with Its Closest Relatives.” *BMC Genomics* 20(1):642. doi: 10.1186/s12864-019-5995-4.

Krysenko, Sergii, Arne Matthews, Tobias Busche, Agnieszka Bera, and Wolfgang Wohlleben. 2021. “Poly- and Monoamine Metabolism in Streptomyces Coelicolor: The New Role of Glutamine Synthetase-Like Enzymes in the Survival under Environmental Stress.” *Microbial Physiology* 1–15. doi: 10.1159/000516644.

Kuypers, Marcel M. M., Hannah K. Marchant, and Boran Kartal. 2018. “The Microbial Nitrogen-Cycling Network.” *Nature Reviews Microbiology* 16(5):263–76. doi: 10.1038/nrmicro.2018.9.

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):12976–83. doi: 10.1021/bi9715371.

Lardi, Martina, Yilei Liu, Gabriela Purtschert, Samanta Bolzan de Campos, and Gabriella Pessi. 2017. “Transcriptome Analysis of Paraburkholderia Phymatum under Nitrogen Starvation and during Symbiosis with Phaseolus Vulgaris.” *Genes* 8(12):389. doi: 10.3390/genes8120389.

Le, Xuyen H., Christopher M. M. Franco, Ross A. Ballard, and Elizabeth A. Drew. 2016. “Isolation and Characterisation of Endophytic Actinobacteria and Their Effect on the Early Growth and Nodulation of Lucerne (Medicago Sativa L.).” *Plant and Soil* 405(1/2):13–24.

Lee, Chi Chung, Yilin Hu, and Markus W. Ribbe. 2009. “Unique Features of the Nitrogenase VFe Protein from Azotobacter Vinelandii.” *Proceedings of the National Academy of Sciences of the United States of America* 106(23):9209–14. doi: 10.1073/pnas.0904408106.

Lehman, L. J., and G. P. Roberts. 1991. “Identification of an Alternative Nitrogenase System in Rhodospirillum Rubrum.” *Journal of Bacteriology* 173(18):5705–11.

Leigh, J. A. 2000. “Nitrogen Fixation in Methanogens: The Archaeal Perspective.” *Current Issues in Molecular Biology* 2(4):125–31.

Letunic, Ivica, and Peer Bork. 2019. “Interactive Tree Of Life (ITOL) v4: Recent Updates and New Developments.” *Nucleic Acids Research* 47(W1):W256–59. doi: 10.1093/nar/gkz239.

Li, Wenyan, Fei Li, Haijun Zeng, Lin Ma, Lingyan Qi, Xiaochun Wang, Wenyin Wang, Zhen Peng, A. Allan Degen, Yanfu Bai, Tao Zhang, Mei Huang, Jin Han, and Zhanhuan Shang. 2021. “Diversity and Variation of Asymbiotic Nitrogen-Fixing Microorganisms in Alpine Grasslands on the Tibetan Plateau.” *Frontiers in Ecology and Evolution* 9:521. doi: 10.3389/fevo.2021.702848.

Li, Xing, Asha Nair, Shengqin Wang, and Liguo Wang. 2015. “Quality Control of RNA-Seq Experiments.” Pp. 137–46 in *RNA Bioinformatics*, *Methods in Molecular Biology*, edited by E. Picardi. New York, NY: Springer.

Liao, Yang, Gordon K. Smyth, and Wei Shi. 2014. “FeatureCounts: An Efficient General Purpose Program for Assigning Sequence Reads to Genomic Features.” *Bioinformatics* 30(7):923–30. doi: 10.1093/bioinformatics/btt656.

Liba, C. m., F. i. s. Ferrara, G. p. Manfio, F. Fantinatti-Garboggini, R. c. Albuquerque, C. Pavan, P. l. Ramos, C. a. Moreira-Filho, and H. r. Barbosa. 2006. “Nitrogen-Fixing Chemo-Organotrophic Bacteria Isolated from Cyanobacteria-Deprived Lichens and Their Ability to Solubilize Phosphate and to Release Amino Acids and Phytohormones.” *Journal of Applied Microbiology* 101(5):1076–86. doi: 10.1111/j.1365-2672.2006.03010.x.

Lin, Yu, Jeffrey Yuan, Mikhail Kolmogorov, Max W. Shen, Mark Chaisson, and Pavel A. Pevzner. 2016. “Assembly of Long Error-Prone Reads Using de Bruijn Graphs.” *Proceedings of the National Academy of Sciences* 113(52):E8396–8405. doi: 10.1073/pnas.1604560113.

Liu, Xiaomeng, Qin Li, Yongbin Li, Guohua Guan, and Sanfeng Chen. 2019. "Paenibacillus Strains with Nitrogen Fixation and Multiple Beneficial Properties for Promoting Plant Growth." *PeerJ* 7. doi: 10.7717/peerj.7445.

Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. "Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2." *Genome Biology* 15(12):550. doi: 10.1186/s13059-014-0550-8.

MacKay, Bruce A., and Michael D. Fryzuk. 2004a. "Dinitrogen Coordination Chemistry: On the Biomimetic Borderlands." *Chemical Reviews* 104(2):385–402. doi: 10.1021/cr020610c.

MacKay, Bruce A., and Michael D. Fryzuk. 2004b. "Dinitrogen Coordination Chemistry: On the Biomimetic Borderlands." *Chemical Reviews* 104(2):385–402. doi: 10.1021/cr020610c.

MacKellar, Drew, Lucas Lieber, Jeffrey S. Norman, Anthony Bolger, Cory Tobin, James W. Murray, Mehtap Oksaksin, Roger L. Chang, Tyler J. Ford, Peter Q. Nguyen, Jimmy Woodward, Hugo R. Permingeat, Neel S. Joshi, Pamela A. Silver, Björn Usadel, Alfred W. Rutherford, Maren L. Friesen, and Jürgen Prell. 2016. "Streptomyces Thermoautotrophicus Does Not Fix Nitrogen." *Scientific Reports* 6. doi: 10.1038/srep20086.

Maia, Luisa B., and José J. G. Moura. 2014. "How Biology Handles Nitrite." *Chemical Reviews* 114(10):5273–5357. doi: 10.1021/cr400518y.

Makhalanyane, Thulani P., Angel Valverde, Eoin Gunnigle, Aline Frossard, Jean-Baptiste Ramond, and Don A. Cowan. 2015. "Microbial Ecology of Hot Desert Edaphic Systems." *FEMS Microbiology Reviews* 39(2):203–21. doi: 10.1093/femsre/fuu011.

Martinez-Argudo, Isabel, Richard Little, Neil Shearer, Philip Johnson, and Ray Dixon. 2004. "The NifL-NifA System: A Multidomain Transcriptional Regulatory Complex That Integrates Environmental Signals." *Journal of Bacteriology* 186(3):601–10. doi: 10.1128/JB.186.3.601-610.2004.

Meher, Prabina K., Tanmaya K. Sahu, Jyotilipsa Mohanty, Shachi Gahoi, Supriya Purru, Monendra Grover, and Atmakuri R. Rao. 2018. "NifPred: Proteome-Wide Identification and Categorization of Nitrogen-Fixation Proteins of Diaztrophs Based on Composition-Transition-Distribution Features Using Support Vector Machine." *Frontiers in Microbiology* 9. doi: 10.3389/fmicb.2018.01100.

Menzel, Peter, Kim Lee Ng, and Anders Krogh. 2016. "Fast and Sensitive Taxonomic Classification for Metagenomics with Kaiju." *Nature Communications* 7(1):11257. doi: 10.1038/ncomms11257.

Mirza, Babur S., and Jorge L. M. Rodrigues. 2012. "Development of a Direct Isolation Procedure for Free-Living Diazotrophs under Controlled Hypoxic Conditions." *Applied and Environmental Microbiology* 78(16):5542–49. doi: 10.1128/AEM.00714-12.

Mistry, Jaina, Sara Chuguransky, Lowri Williams, Matloob Qureshi, Gustavo A. Salazar, Erik L. L. Sonnhammer, Silvio C. E. Tosatto, Lisanna Paladin, Shriya Raj, Lorna J. Richardson, Robert D. Finn, and Alex Bateman. 2021. "Pfam: The Protein Families Database in 2021." *Nucleic Acids Research* 49(D1):D412–19. doi: 10.1093/nar/gkaa913.

Moore, A. W., and J. H. Becking. 1963. "Nitrogen Fixation by Bacillus Strains Isolated from Nigerian Soils." *Nature* 198(4883):915–16. doi: 10.1038/198915a0.

Moran, Mary Ann, Brandon Satinsky, Scott M. Gifford, Haiwei Luo, Adam Rivers, Leong-Keat Chan, Jun Meng, Bryndan P. Durham, Chen Shen, Vanessa A. Varaljay, Christa B. Smith, Patricia L. Yager, and Brian M. Hopkinson. 2013. "Sizing up Metatranscriptomics." *The ISME Journal* 7(2):237–43. doi: 10.1038/ismej.2012.94.

Moreno-Vivián, C., P. Cabello, M. Martínez-Luque, R. Blasco, and F. Castillo. 1999. "Prokaryotic Nitrate Reduction: Molecular Properties and Functional Distinction among Bacterial Nitrate Reductases." *Journal of Bacteriology* 181(21):6573–84. doi: 10.1128/JB.181.21.6573-6584.1999.

Muñoz-Marín, María del Carmen, Irina N. Shilova, Tuo Shi, Hanna Farnelid, Ana María Cabello, and Jonathan P. Zehr. 2019. "The Transcriptional Cycle Is Suited to Daytime N2 Fixation in the Unicellular Cyanobacterium 'Candidatus Atelocyanobacterium Thalassa' (UCYN-A)." *MBio* 10(1). doi: 10.1128/mBio.02495-18.

Muyzer, G., E. C. de Waal, and A. G. Uitterlinden. 1993. "Profiling of Complex Microbial Populations by Denaturing Gradient Gel Electrophoresis Analysis of Polymerase Chain Reaction-Amplified Genes Coding for 16S RRNA." *Applied and Environmental Microbiology* 59(3):695–700. doi: 10.1128/aem.59.3.695-700.1993.

Muyzer, G, E. C. de Waal, and A. G. Uitterlinden. 1993. "Profiling of Complex Microbial Populations by Denaturing Gradient Gel Electrophoresis Analysis of Polymerase Chain Reaction-Amplified Genes Coding for 16S RRNA." *Applied and Environmental Microbiology* 59(3):695–700.

Navarro-González, Rafael, Christopher P. McKay, and Delphine Nna Mvondo. 2001. "A Possible Nitrogen Crisis for Archaean Life Due to Reduced Nitrogen Fixation by Lightning." *Nature* 412(6842):61–64. doi: 10.1038/35083537.

Navarro-Noya, Yendi E., Esperanza Martínez-Romero, and César Hernández-Rodríguez. 2013. "Potential Plant-Growth-Promoting and Nitrogen-Fixing Bacteria Associated with Pioneer Plants Growing on Mine Tailings." in *Molecular Microbial Ecology of the Rhizosphere*.

Nelson, Michaeline B., Adam C. Martiny, and Jennifer B. H. Martiny. 2016. "Global Biogeography of Microbial Nitrogen-Cycling Traits in Soil." *Proceedings of the National Academy of Sciences* 113(29):8033–40. doi: 10.1073/pnas.1601070113.

Neufeld, Josh D., Jyotsna Vohra, Marc G. Dumont, Tillmann Lueders, Mike Manefield, Michael W. Friedrich, and J. Colin Murrell. 2007. "DNA Stable-Isotope Probing." *Nature Protocols* 2(4):860–66. doi: 10.1038/nprot.2007.109.

Newton, William E. 2015. "Recent Advances in Understanding Nitrogenases and How They Work." Pp. 5–20 in *Biological Nitrogen Fixation*. John Wiley & Sons, Ltd.

Nindita, Yosi, Zhisheng Cao, Amirudin Akhmad Fauzi, Aiko Teshima, Yuya Misaki, Rukman Muslimin, Yingjie Yang, Yuh Shiwa, Hirofumi Yoshikawa, Michihira Tagami, Alexander Lezhava, Jun Ishikawa, Makoto Kuroda, Tsuyoshi Sekizuka, Kuninobu Inada, Haruyasu Kinashi, and Kenji Arakawa. 2019. "The Genome Sequence of Streptomyces Rochei 7434AN4, Which Carries a Linear Chromosome and Three Characteristic Linear Plasmids." *Scientific Reports* 9(1):10973. doi: 10.1038/s41598-019-47406-y.

Ninfa, A. J., P. Jiang, M. R. Atkinson, and J. A. Peliska. 2000. "Integration of Antagonistic Signals in the Regulation of Nitrogen Assimilation in Escherichia Coli." *Current Topics in Cellular Regulation* 36:31–75. doi: 10.1016/s0070-2137(01)80002-9.

Normand, P., S. Orso, Benoit Cournoyer, P. Jeannin, C. Chapelon, Jeffrey Dawson, Lyudmila Evtushenko, and Arvind Misra. 1996. "Molecular Phylogeny of the Genus Frankia and Related Genera and Emendation of the Family Frankiaceae." *International Journal of Systematic Bacteriology* 46:1–9. doi: 10.1099/00207713-46-1-1.

Ogawa, K., E. Akagawa, K. Yamane, Z. W. Sun, M. LaCelle, P. Zuber, and M. M. Nakano. 1995. "The NasB Operon and NasA Gene Are Required for Nitrate/Nitrite Assimilation in Bacillus Subtilis." *Journal of Bacteriology* 177(5):1409–13. doi: 10.1128/jb.177.5.1409-1413.1995.

Oshiki, Mamoru, Takahiro Segawa, and Satoshi Ishii. 2018. "Nitrogen Cycle Evaluation (NiCE) Chip for Simultaneous Analysis of Multiple N Cycle-Associated Genes." *Applied and Environmental Microbiology* 84(8):e02615-17. doi: 10.1128/AEM.02615-17.

Pajares, Silvia, and Brendan J. M. Bohannan. 2016. "Ecology of Nitrogen Fixing, Nitrifying, and Denitrifying Microorganisms in Tropical Forest Soils." *Frontiers in Microbiology* 7:1045. doi: 10.3389/fmicb.2016.01045.

Park, Myoungsu, Chungwoo Kim, Jinchul Yang, Hyoungseok Lee, Wansik Shin, Seunghwan Kim, and Tongmin Sa. 2005. "Isolation and Characterization of Diazotrophic Growth Promoting Bacteria from Rhizosphere of Agricultural Crops of Korea." *Microbiological Research* 160(2):127–33. doi: 10.1016/j.micres.2004.10.003.

Pastorino, Graciela Noemí, Virginia Martínez Alcántara, Ismael Malbrán, L. Videira, J. Sarinelli, and Pedro Alberto Balatti. 2015. "Ensifer (Sinorhizobium) Fredii Interacted More Efficiently than Bradyrhizobium Japonicum with Soybean." doi: 10.9734/JAERI/2015/13163.

Pathani, Neemisha, S. K. Gosal, G. S. Saroa, and Yogesh Vikal. 2014. "Molecular Characterization of Diazotrophic Bacteria Isolated from Rhizosphere of Wheat Cropping System from Central Plain Region of Punjab." *African Journal of Microbiology Research* 8(9):862–71. doi: 10.5897/AJMR2013.5948.

Patra, Ashok K., Luc Abbadie, Annie Clays-Josserand, Valérie Degrange, Susan J. Grayston, Nadine Guillaumaud, Pierre Loiseau, Frédérique Louault, Shahid Mahmood, Sylvie Nazaret, Laurent Philippot, Franck Poly, James I. Prosser, and Xavier Le Roux. 2006. "Effects of Management Regime and Plant Species on the Enzyme Activity and Genetic Structure of N-Fixing, Denitrifying and Nitrifying Bacterial Communities in Grassland Soils." *Environmental Microbiology* 8(6):1005–16. doi: 10.1111/j.1462-2920.2006.00992.x.

Patro, Rob, Geet Duggal, Michael I. Love, Rafael A. Irizarry, and Carl Kingsford. 2017. "Salmon: Fast and Bias-Aware Quantification of Transcript Expression Using Dual-Phase Inference." *Nature Methods* 14(4):417–19. doi: 10.1038/nmeth.4197.

Paulot, F., D. J. Jacob, M. T. Johnson, T. G. Bell, A. R. Baker, W. C. Keene, I. D. Lima, S. C. Doney, and C. A. Stock. 2015. "Global Oceanic Emission of Ammonia: Constraints from Seawater and Atmospheric Observations." *Global Biogeochemical Cycles* 29(8):1165–78. doi: 10.1002/2015gb005106.

Poly, F., L. J. Monrozier, and R. Bally. 2001. "Improvement in the RFLP Procedure for Studying the Diversity of NifH Genes in Communities of Nitrogen Fixers in Soil." *Research in Microbiology* 152(1):95–103. doi: 10.1016/s0923-2508(00)01172-4.

Postgate, John Raymond, William Duncan Paterson Stewart, and T. Rosswall. 1982. "Biological Nitrogen Fixation: Fundamentals." *Philosophical Transactions of the Royal Society of London. B, Biological Sciences* 296(1082):375–85. doi: 10.1098/rstb.1982.0013.

Poza-Carrión, César, Carlos Echavarri-Erasun, and Luis M. Rubio. 2015a. “Regulation of Nif Gene Expression in Azotobacter Vinelandii.” Pp. 99–108 in *Biological Nitrogen Fixation*. John Wiley & Sons, Ltd.

Poza-Carrión, César, Carlos Echavarri-Erasun, and Luis M. Rubio. 2015b. “Regulation of Nif Gene Expression in Azotobacter Vinelandii.” Pp. 99–108 in *Biological Nitrogen Fixation*. John Wiley & Sons, Ltd.

Price, Morgan N., Paramvir S. Dehal, and Adam P. Arkin. 2010. “FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments.” *PLOS ONE* 5(3):e9490. doi: 10.1371/journal.pone.0009490.

Proctor, Michael H., and P. W. Wilson. 1959. “Nitrogen Fixation by Achromobacter Spp.” *Archiv Für Mikrobiologie* 32(3):254–60. doi: 10.1007/BF00409346.

Raittz, Roberto Tadeu, Camilla Reginatto De Pierri, Marta Maluk, Marcelo Bueno Batista, Manuel Carmona, Madan Junghare, Helisson Faoro, Leonardo M. Cruz, Federico Battistoni, Emanuel de Souza, Fábio de Oliveira Pedrosa, Wen-Ming Chen, Philip S. Poole, Ray A. Dixon, and Euan K. James. 2021. “Comparative Genomics Provides Insights into the Taxonomy of Azoarcus and Reveals Separate Origins of Nif Genes in the Proposed Azoarcus and Aromatoleum Genera.” *Genes* 12(1):71. doi: 10.3390/genes12010071.

Rambaut, Andrew. 2012. *FigTree v1. 4*.

Ramos, Patrícia L., Stefanie Van Trappen, Fabiano L. Thompson, Rafael C. S. Rocha, Heloiza R. Barbosa, Paul De Vos, and Carlos A. Moreira-Filho. 2011. “Screening for Endophytic Nitrogen-Fixing Bacteria in Brazilian Sugar Cane Varieties Used in Organic Farming and Description of Stenotrophomonas Pavanii Sp. Nov.” *International Journal of Systematic and Evolutionary Microbiology* 61(Pt 4):926–31. doi: 10.1099/ijs.0.019372-0.

Rascio, N., and N. La Rocca. 2008. “Biological Nitrogen Fixation.” Pp. 412–19 in *Encyclopedia of Ecology*, edited by S. E. Jørgensen and B. D. Fath. Oxford: Academic Press.

Raymond, Jason, Janet L. Siefert, Christopher R. Staples, and Robert E. Blankenship. 2004. “The Natural History of Nitrogen Fixation.” *Molecular Biology and Evolution* 21(3):541–54. doi: 10.1093/molbev/msh047.

Rehder, D. 2000. “Vanadium Nitrogenase.” *Journal of Inorganic Biochemistry* 80(1–2):133–36. doi: 10.1016/s0162-0134(00)00049-0.

Ren, Min, Zhufeng Zhang, Xuelian Wang, Zhiwei Zhou, Dong Chen, Hui Zeng, Shumiao Zhao, Lingling Chen, Yuanliang Hu, Changyi Zhang, Yunxiang Liang, Qunxin She, Yi Zhang, and Nan Peng. 2018. “Diversity and Contributions to Nitrogen

Cycling and Carbon Fixation of Soil Salinity Shaped Microbial Communities in Tarim Basin." *Frontiers in Microbiology* 9:431. doi: 10.3389/fmicb.2018.00431.

Revell, Liam J., and Scott A. Chamberlain. 2014. "Rphylip: An R Interface for PHYLIP." *Methods in Ecology and Evolution* 5(9):976–81. doi: https://doi.org/10.1111/2041-210X.12233.

Ribbe, Markus W., ed. 2011. *Nitrogen Fixation: Methods and Protocols*. Humana Press.

Rilling, Joaquin I., Jacquelinne J. Acuña, Michael J. Sadowsky, and Milko A. Jorquera. 2018. "Putative Nitrogen-Fixing Bacteria Associated With the Rhizosphere and Root Endosphere of Wheat Plants Grown in an Andisol From Southern Chile." *Frontiers in Microbiology* 9. doi: 10.3389/fmicb.2018.02710.

Risser, Paul G. 1985. "Grasslands." Pp. 232–56 in *Physiological Ecology of North American Plant Communities*, edited by B. F. Chabot and H. A. Mooney. Dordrecht: Springer Netherlands.

RoÂzÇycki, H., H. Dahm, E. Strzelczyk, and C. Y. Li. 1999. "Diazotrophic Bacteria in Root-Free Soil and in the Root Zone of Pine (Pinus Sylvestris L.) and Oak (Quercus Robur L.)." *Applied Soil Ecology* 12.

Roberts, GARY P., TANYA MAcNEIL, DOUGLAS MAcNEIL, and WINSTON J. Brill. 1978. "Regulation and Characterization of Protein Products Coded by the Nif (Nitrogen Fixation) Genes of Klebsiella Pneumoniae." *Journal of Bacteriology* 136(1):267–79.

Rodríguez-Quiñones, F., R. Bosch, and J. Imperial. 1993. "Expression of the NifBfdxNnifOQ Region of Azotobacter Vinelandii and Its Role in Nitrogenase Activity." *Journal of Bacteriology* 175(10):2926–35. doi: 10.1128/jb.175.10.2926-2935.1993.

Rosenberg, Eugene, Edward F. DeLong, Erko Stackebrandt, Stephen Lory, and Fabiano Thompson, eds. 2013. *The Prokaryotes: Prokaryotic Physiology and Biochemistry*. 4th ed. Berlin Heidelberg: Springer-Verlag.

Rubio, Luis M., and Paul W. Ludden. 2005. "Maturation of Nitrogenase: A Biochemical Puzzle." *Journal of Bacteriology* 187(2):405–14. doi: 10.1128/JB.187.2.405-414.2005.

Rubio, Luis M., and Paul W. Ludden. 2008. "Biosynthesis of the Iron-Molybdenum Cofactor of Nitrogenase." *Annual Review of Microbiology* 62:93–111. doi: 10.1146/annurev.micro.62.081307.162737.

Saleh-Lakha, Saleema, Michelle Miller, Rachel G. Campbell, Kim Schneider, Parastu Elahimanesh, Miranda M. Hart, and Jack T. Trevors. 2005. "Microbial Gene

Expression in Soil: Methods, Applications and Challenges." *Journal of Microbiological Methods* 63(1):1–19. doi: 10.1016/j.mimet.2005.03.007.

Sathya, Arumugam, Rajendran Vijayabharathi, Vadlamudi Srinivas, and Subramaniam Gopalakrishnan. 2016. "Plant Growth-Promoting Actinobacteria on Chickpea Seed Mineral Density: An Upcoming Complementary Tool for Sustainable Biofortification Strategy." *3 Biotech* 6(2):138. doi: 10.1007/s13205-016-0458-y.

Savcı, Serpil. 2012. "An Agricultural Pollutant: Chemical Fertilizer." *International Journal of Environmental Science and Development* 3. doi: 10.7763/IJESD.2012.V3.191.

Schneider, K., U. Gollan, M. Dröttboom, S. Selsemeier-Voigt, and A. Müller. 1997. "Comparative Biochemical Characterization of the Iron-Only Nitrogenase and the Molybdenum Nitrogenase from Rhodobacter Capsulatus." *European Journal of Biochemistry* 244(3):789–800. doi: 10.1111/j.1432-1033.1997.t01-1-00789.x.

Schneider, K., A. Müller, U. Schramm, and W. Klipp. 1991. "Demonstration of a Molybdenum- and Vanadium-Independent Nitrogenase in a NifHDK-Deletion Mutant of Rhodobacter Capsulatus." *European Journal of Biochemistry* 195(3):653–61. doi: 10.1111/j.1432-1033.1991.tb15750.x.

Schütte, Ursel M. E., Zaid Abdo, Stephen J. Bent, Conrad Shyu, Christopher J. Williams, Jacob D. Pierson, and Larry J. Forney. 2008. "Advances in the Use of Terminal Restriction Fragment Length Polymorphism (T-RFLP) Analysis of 16S RRNA Genes to Characterize Microbial Communities." *Applied Microbiology and Biotechnology* 80(3):365–80. doi: 10.1007/s00253-008-1565-4.

Seemann, Torsten. 2014. "Prokka: Rapid Prokaryotic Genome Annotation." *Bioinformatics* 30(14):2068–69. doi: 10.1093/bioinformatics/btu153.

Sellstedt, Anita, and Kerstin H. Richau. 2013. "Aspects of Nitrogen-Fixing Actinobacteria, in Particular Free-Living and Symbiotic Frankia." *FEMS Microbiology Letters* 342(2):179–86. doi: 10.1111/1574-6968.12116.

Setubal, João C., Patricia dos Santos, Barry S. Goldman, Helga Ertesvåg, Guadelupe Espin, Luis M. Rubio, Svein Valla, Nalvo F. Almeida, Divya Balasubramanian, Lindsey Cromes, Leonardo Curatti, Zijin Du, Eric Godsy, Brad Goodner, Kaitlyn Hellner-Burris, José A. Hernandez, Katherine Houmiel, Juan Imperial, Christina Kennedy, Timothy J. Larson, Phil Latreille, Lauren S. Ligon, Jing Lu, Mali Mærk, Nancy M. Miller, Stacie Norton, Ina P. O'Carroll, Ian Paulsen, Estella C. Raulfs, Rebecca Roemer, James Rosser, Daniel Segura, Steve Slater, Shawn L. Stricklin, David J. Studholme, Jian Sun, Carlos J. Viana, Erik Wallin, Baomin Wang, Cathy Wheeler, Huijun Zhu, Dennis R. Dean, Ray Dixon, and Derek Wood. 2009. "Genome Sequence of Azotobacter Vinelandii, an Obligate Aerobe

Specialized To Support Diverse Anaerobic Metabolic Processes." *Journal of Bacteriology* 191(14):4534–45. doi: 10.1128/JB.00504-09.

Shendure, Jay, and Hanlee Ji. 2008. "Next-Generation DNA Sequencing." *Nature Biotechnology* 26(10):1135–45. doi: 10.1038/nbt1486.

Shoun, Hirofumi, Shinya Fushinobu, Li Jiang, Sang-Wan Kim, and Takayoshi Wakagi. 2012. "Fungal Denitrification and Nitric Oxide Reductase Cytochrome P450nor." *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 367(1593):1186–94. doi: 10.1098/rstb.2011.0335.

Simonsen, Martin, Thomas Mailund, and Christian N. S. Pedersen. 2008. "Rapid Neighbour-Joining." Pp. 113–22 in *Algorithms in Bioinformatics*, *Lecture Notes in Computer Science*, edited by K. A. Crandall and J. Lagergren. Berlin, Heidelberg: Springer.

Simpson, Frank B., and Robert H. Burris. 1984. "A Nitrogen Pressure of 50 Atmospheres Does Not Prevent Evolution of Hydrogen by Nitrogenase." *Science* 224(4653):1095–97. doi: 10.1126/science.6585956.

Smith, Barry E. 2002. "Nitrogenase Reveals Its Inner Secrets." *Science* 297(5587):1654–55. doi: 10.1126/science.1076659.

Speth, Daan, Baolan Hu, Niek Bosch, Jan Keltjens, Hendrik Stunnenberg, and Mike Jetten. 2012. "Comparative Genomics of Two Independently Enriched 'Candidatus Kuenenia Stuttgartiensis' Anammox Bacteria." *Frontiers in Microbiology* 3:307. doi: 10.3389/fmicb.2012.00307.

Spurgeon, Sandra L., Robert C. Jones, and Ramesh Ramakrishnan. 2008. "High Throughput Gene Expression Measurement with Real Time PCR in a Microfluidic Dynamic Array." *PLOS ONE* 3(2):e1662. doi: 10.1371/journal.pone.0001662.

Stal, Lucas J. 2015. "Nitrogen Fixation in Cyanobacteria." Pp. 1–9 in *eLS*. American Cancer Society.

Stein, Lisa Y., and Martin G. Klotz. 2016. "The Nitrogen Cycle." *Current Biology: CB* 26(3):R94-98. doi: 10.1016/j.cub.2015.12.021.

Steven, Blaire, Cheryl R. Kuske, La Verne Gallegos-Graves, Sasha C. Reed, and Jayne Belnap. 2015. "Climate Change and Physical Disturbance Manipulations Result in Distinct Biological Soil Crust Communities." *Applied and Environmental Microbiology* 81(21):7448–59. doi: 10.1128/AEM.01443-15.

Stüeken, Eva E., Roger Buick, Bradley M. Guy, and Matthew C. Koehler. 2015. "Isotopic Evidence for Biological Nitrogen Fixation by Molybdenum-Nitrogenase from 3.2 Gyr." *Nature* 520(7549):666–69. doi: 10.1038/nature14180.

Su, Can, Liping Lei, Yanqing Duan, Ke-Qin Zhang, and Jinkui Yang. 2012. "Culture-Independent Methods for Studying Environmental Microorganisms: Methods, Application, and Perspective." *Applied Microbiology and Biotechnology* 93(3):993–1003. doi: 10.1007/s00253-011-3800-7.

Sutton, M. A., A. Bleeker, C. M. Howard, M. Bekunda, B. Grizzetti, W. de Vries, H. J. M. van Grinsven, Y. P. Abrol, T. K. Adhya, G. Billen, E. A. Davidson, A. Datta, R. Diaz, J. W. Erisman, X. J. Liu, O. Oenema, C. Palm, N. Raghuram, S. Reis, R. W. Scholz, T. Sims, H. Westhoek, and F. S. Zhang. 2013. "Our Nutrient World: The Challenge to Produce More Food and Energy with Less Pollution."

Taylor, Benton N., Robin L. Chazdon, and Duncan N. L. Menge. 2019. "Successional Dynamics of Nitrogen Fixation and Forest Growth in Regenerating Costa Rican Rainforests." *Ecology* 100(4):e02637. doi: 10.1002/ecy.2637.

Tezcan, F. Akif, Jens T. Kaiser, Debarshi Mustafi, Mika Y. Walton, James B. Howard, and Douglas C. Rees. 2005. "Nitrogenase Complexes: Multiple Docking Sites for a Nucleotide Switch Protein." *Science (New York, N.Y.)* 309(5739):1377–80. doi: 10.1126/science.1115653.

Thorneley, R. N. F., and D. J. Lowe. 1996. "Nitrogenase: Substrate Binding and Activation." *JBIC Journal of Biological Inorganic Chemistry* 1(6):576–80. doi: 10.1007/s007750050095.

Tosun, İsmail. 2012. "Ammonium Removal from Aqueous Solutions by Clinoptilolite: Determination of Isotherm and Thermodynamic Parameters and Comparison of Kinetics by the Double Exponential Model and Conventional Kinetic Models." *International Journal of Environmental Research and Public Health* 9(3):970–84. doi: 10.3390/ijerph9030970.

Trujillo, Martha E., Raúl Riesco, Patricia Benito, and Lorena Carro. 2015. "Endophytic Actinobacteria and the Interaction of Micromonospora and Nitrogen Fixing Plants." *Frontiers in Microbiology* 6:1341. doi: 10.3389/fmicb.2015.01341.

Tsementzi, Despina, Jieying Wu, Samuel Deutsch, Sangeeta Nath, Luis M. Rodriguez-R, Andrew S. Burns, Piyush Ranjan, Neha Sarode, Rex R. Malmstrom, Cory C. Padilla, Benjamin K. Stone, Laura A. Bristow, Morten Larsen, Jennifer B. Glass, Bo Thamdrup, Tanja Woyke, Konstantinos T. Konstantinidis, and Frank J. Stewart. 2016. "SAR11 Bacteria Linked to Ocean Anoxia and Nitrogen Loss." *Nature* 536(7615):179–83. doi: 10.1038/nature19068.

Tu, Qichao, Ye Deng, Qingyun Yan, Lina Shen, Lu Lin, Zhili He, Liyou Wu, Joy Van Nostrand, Vanessa Buzzard, Sean Michaletz, Brian Enquist, Michael Weiser, Michael Kaspari, Robert Waide, James Brown, and Jizhong Zhou. 2016. "Biogeographic Patterns of Soil Diazotrophic Communities across Six Forests in the North America." *Molecular Ecology* 25. doi: 10.1111/mec.13651.

Ueda, T., Y. Suga, N. Yahiro, and T. Matsuguchi. 1995. “Remarkable N2-Fixing Bacterial Diversity Detected in Rice Roots by Molecular Evolutionary Analysis of NifH Gene Sequences.” *Journal of Bacteriology* 177(5):1414–17. doi: 10.1128/jb.177.5.1414-1417.1995.

Unkovich, M., and J. Baldock. 2008. “Measurement of Asymbiotic N2 Fixation in Australian Agriculture.” doi: 10.1016/J.SOILBIO.2008.08.021.

Upchurch, R. G., and L. E. Mortenson. 1980. “In Vivo Energetics and Control of Nitrogen Fixation: Changes in the Adenylate Energy Charge and Adenosine 5’-Diphosphate/Adenosine 5’-Triphosphate Ratio of Cells during Growth on Dinitrogen versus Growth on Ammonia.” *Journal of Bacteriology* 143(1):274–84.

Van Dommelen, Anne, and Jos Vanderleyden. 2007. “Associative Nitrogen Fixation.” Pp. 179–92 in *Biology of the Nitrogen Cycle*.

Vance, Carroll P. 2001. “Symbiotic Nitrogen Fixation and Phosphorus Acquisition. Plant Nutrition in a World of Declining Renewable Resources.” *Plant Physiology* 127(2):390–97. doi: 10.1104/pp.010331.

Venter, J. Craig, Karin Remington, John F. Heidelberg, Aaron L. Halpern, Doug Rusch, Jonathan A. Eisen, Dongying Wu, Ian Paulsen, Karen E. Nelson, William Nelson, Derrick E. Fouts, Samuel Levy, Anthony H. Knap, Michael W. Lomas, Ken Nealson, Owen White, Jeremy Peterson, Jeff Hoffman, Rachel Parsons, Holly Baden-Tillson, Cynthia Pfannkoch, Yu-Hui Rogers, and Hamilton O. Smith. 2004. “Environmental Genome Shotgun Sequencing of the Sargasso Sea.” *Science* 304(5667):66–74. doi: 10.1126/science.1093857.

Ventura, Marco, Carlos Canchaya, Andreas Tauch, Govind Chandra, Gerald F. Fitzgerald, Keith F. Chater, and Douwe van Sinderen. 2007. “Genomics of Actinobacteria: Tracing the Evolutionary History of an Ancient Phylum.” *Microbiology and Molecular Biology Reviews : MMBR* 71(3):495–548. doi: 10.1128/MMBR.00005-07.

Vitousek, Peter M., Duncan N. L. Menge, Sasha C. Reed, and Cory C. Cleveland. 2013. “Biological Nitrogen Fixation: Rates, Patterns and Ecological Controls in Terrestrial Ecosystems.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 368(1621):20130119. doi: 10.1098/rstb.2013.0119.

Vos, Paul, George Garrity, Dorothy Jones, Noel R. Krieg, Wolfgang Ludwig, Fred A. Rainey, Karl-Heinz Schleifer, and William B. Whitman. 2011. *Bergey’s Manual of Systematic Bacteriology: Volume 3: The Firmicutes*. Springer Science & Business Media.

Wang, Chao, Manman Zheng, Wenfeng Song, Shilin Wen, Boren Wang, Chunquan Zhu, and Renfang Shen. 2017. “Impact of 25 Years of Inorganic Fertilization on Diazotrophic Abundance and Community Structure in an Acidic Soil in Southern

China." *Soil Biology and Biochemistry* 113:240–49. doi: 10.1016/j.soilbio.2017.06.019.

Weaver, J. 1954. "North American Prairie." *Papers of John E. Weaver (1884-1956)*.

Weisburg, W. G., S. M. Barns, D. A. Pelletier, and D. J. Lane. 1991. "16S Ribosomal DNA Amplification for Phylogenetic Study." *Journal of Bacteriology* 173(2):697–703. doi: 10.1128/jb.173.2.697-703.1991.

Weiss, Madeline C., Filipa L. Sousa, Natalia Mrnjavac, Sinje Neukirchen, Mayo Roettger, Shijulal Nelson-Sathi, and William F. Martin. 2016. "The Physiology and Habitat of the Last Universal Common Ancestor." *Nature Microbiology* 1(9):1–8. doi: 10.1038/nmicrobiol.2016.116.

Wertz, John T., Eunji Kim, John A. Breznak, Thomas M. Schmidt, and Jorge L. M. Rodrigues. 2012. "Genomic and Physiological Characterization of the Verrucomicrobia Isolate Diplosphaera Colitermitum Gen. Nov., Sp. Nov., Reveals Microaerophily and Nitrogen Fixation Genes." *Applied and Environmental Microbiology* 78(5):1544–55. doi: 10.1128/AEM.06466-11.

Wouters, Johanna, John A. Raven, Susanna Minnhagen, and Sven Janson. 2009. "The Luggage Hypothesis: Comparisons of Two Phototrophic Hosts with Nitrogen-Fixing Cyanobacteria and Implications for Analogous Life Strategies for Kleptoplastids/Secondary Symbiosis in Dinoflagellates." *Symbiosis* 49(2):61–70. doi: 10.1007/s13199-009-0020-3.

Wu, Hao, Guiguang Chen, Yaxi Bian, Wei Zeng, Bihong Sun, and Zhiqun Liang. 2017. "Identification and Characterization of a New Agar-Degrading Strain with the Novel Properties of Saccharides Inhibition and Nitrogen Fixation." *Journal of Microbiology* 55(6):475–82. doi: 10.1007/s12275-017-6464-x.

Xiao, Zhenggao, Sergio Rasmann, Le Yue, Fei Lian, Hua Zou, and Zhenyu Wang. 2019. "The Effect of Biochar Amendment on N-Cycling Genes in Soils: A Meta-Analysis." *Science of The Total Environment* 696:133984. doi: 10.1016/j.scitotenv.2019.133984.

Yamada, Akinori, T. Inoue, S. Noda, Yuichi Hongoh, and M. Ohkuma. 2007. "Evolutionary Trend of Phylogenetic Diversity of Nitrogen Fixation Genes in the Gut Community of Wood-Feeding Termites." *Molecular Ecology* 16:3768–77. doi: 10.1111/j.1365-294X.2007.03326.x.

Yang, Zhi-Yong, Dennis R. Dean, and Lance C. Seefeldt. 2011. "Molybdenum Nitrogenase Catalyzes the Reduction and Coupling of CO to Form Hydrocarbons♦." *The Journal of Biological Chemistry* 286(22):19417–21. doi: 10.1074/jbc.M111.229344.

Yoshida, Nobuyuki, Satoshi Inaba, and Hiroshi Takagi. 2014. “Utilization of Atmospheric Ammonia by an Extremely Oligotrophic Bacterium, Rhodococcus Erythropolis N9T-4.” *Journal of Bioscience and Bioengineering* 117(1):28–32. doi: 10.1016/j.jbiosc.2013.06.005.

Young, J. P. W. 2005. “The Phylogeny and Evolution of Nitrogenases.” Pp. 221–41 in *Genomes and Genomics of Nitrogen-fixing Organisms*, *Nitrogen Fixation: Origins, Applications, and Research Progress*, edited by R. Palacios and W. E. Newton. Dordrecht: Springer Netherlands.

Yousuf, Jesmi, Jabir Thajudeen, Mujeeb Rahiman, Soumya Krishnankutty, Aneesa P. Alikunj, and Mohamed H. A. Abdulla. 2017. “Nitrogen Fixing Potential of Various Heterotrophic Bacillus Strains from a Tropical Estuary and Adjacent Coastal Regions.” *Journal of Basic Microbiology* 57(11):922–32. doi: 10.1002/jobm.201700072.

Zehr, Jonathan P., and Douglas G. Capone. 2020. “Changing Perspectives in Marine Nitrogen Fixation.” *Science* 368(6492):eaay9514. doi: 10.1126/science.aay9514.

Zehr, Jonathan P., Bethany D. Jenkins, Steven M. Short, and Grieg F. Steward. 2003. “Nitrogenase Gene Diversity and Microbial Community Structure: A Cross-System Comparison.” *Environmental Microbiology* 5(7):539–54. doi: 10.1046/j.1462-2920.2003.00451.x.

Zehr, Jonathan P., Irina N. Shilova, Hanna M. Farnelid, Maria Del Carmen Muñoz-Marín, and Kendra A. Turk-Kubo. 2016. “Unusual Marine Unicellular Symbiosis with the Nitrogen-Fixing Cyanobacterium UCYN-A.” *Nature Microbiology* 2:16214. doi: 10.1038/nmicrobiol.2016.214.

Zhang, Chao, Maryam Rabiee, Erfan Sayyari, and Siavash Mirarab. 2018. “ASTRAL-III: Polynomial Time Species Tree Reconstruction from Partially Resolved Gene Trees.” *BMC Bioinformatics* 19(6):153. doi: 10.1186/s12859-018-2129-y.

Zhang, Xinning, Bess B. Ward, and Daniel M. Sigman. 2020. “Global Nitrogen Cycle: Critical Enzymes, Organisms, and Processes for Nitrogen Budgets and Dynamics.” *Chemical Reviews* 120(12):5308–51. doi: 10.1021/acs.chemrev.9b00613.

Zheng, Hao, Carsten Dietrich, Renate Radek, and Andreas Brune. 2016. “Endomicrobium Proavitum, the First Isolate of Endomicrobia Class. Nov. (Phylum Elusimicrobia)--an Ultramicrobacterium with an Unusual Cell Cycle That Fixes Nitrogen with a Group IV Nitrogenase.” *Environmental Microbiology* 18(1):191–204. doi: 10.1111/1462-2920.12960.

Zhu, Baoli, Jiaqi Wang, Lauren M. Bradford, Katharina Ettwig, Baolan Hu, and Tillmann Lueders. 2019. “Nitric Oxide Dismutase (Nod) Genes as a Functional Marker for

the Diversity and Phylogeny of Methane-Driven Oxygenic Denitrifiers.” *Frontiers in Microbiology* 10:1577. doi: 10.3389/fmicb.2019.01577.

Zhu, Qiyun, Uyen Mai, Wayne Pfeiffer, Stefan Janssen, Francesco Asnicar, Jon G. Sanders, Pedro Belda-Ferre, Gabriel A. Al-Ghalith, Evguenia Kopylova, Daniel McDonald, Tomasz Kosciolek, John B. Yin, Shi Huang, Nimaichand Salam, Jian-Yu Jiao, Zijun Wu, Zhenjiang Z. Xu, Kalen Cantrell, Yimeng Yang, Erfan Sayyari, Maryam Rabiee, James T. Morton, Sheila Podell, Dan Knights, Wen-Jun Li, Curtis Huttenhower, Nicola Segata, Larry Smarr, Siavash Mirarab, and Rob Knight. 2019. “Phylogenomics of 10,575 Genomes Reveals Evolutionary Proximity between Domains Bacteria and Archaea.” *Nature Communications* 10(1):5477. doi: 10.1038/s41467-019-13443-4.

Appendix

Appendix 1 Genera isolated in this study and their status regarding diazotrophy.

Genus	Isolated In	Total	NifH	Comment on Diazotrophy	Reference
Streptomyces	Both	104	+ (30)	Reported and rejected	(Gadkari et al. 1992; MacKellar et al. 2016)
Pseudomonas	Both	52	+ (1)	Yes	
Rhizobium	Both	39	+ (6)	Yes	
Microbacterium	Both	27	+ (12)	Reported but no genomic evidence	(Gtari et al. 2012; Rilling et al. 2018)
Rhodococcus	Both	27	+ (4)	Oligotrophic	(Blasco et al. 2001; Joshi et al. 2019b; Yoshida et al. 2014)
Bacillus	Both	25	+ (1)	Reported	(Moore and Becking 1963)
Stenotrophomonas	Both	25	+ (4)	Yes	(Ramos et al. 2011)
Phyllobacterium	Both	24	+ (2)	Yes	(Ikbal et al. 2020)
Kribbella	Both	11	-		
Variovorax	Both	10	-	Genomic evidence	Chapter 1
Flavobacterium	Both	8	+ (1)	Yes	(Kämpfer et al. 2015)
Inquilinus	Both	6	-	No	(Navarro-Noya, Martínez-Romero, and Hernández-Rodríguez 2013)
Ensifer	Both	5	-	Yes	(Pastorino et al. 2015)
Kitasatospora	Both	5	-	Growth promoter	(Trujillo et al. 2015)
Mesorhizobium	Both	4	+ (2)	Yes	
Nocardioides	Both	4	+ (2)	Growth promoter	(Trujillo et al. 2015)
Paraburkholderia	Both	4	-	Yes	(Lardi et al. 2017)
Bosea	Both	3	-	Related to rhizobia	(Rilling et al. 2018)
Cellulomonas	Aerobic	3	-	Reported co-culture with diazotrophs	(Halsall and Gibson 1985)
Duganella	Both	3	-	Plant beneficial	(Fang et al. 2019)
Mycobacterium	Aerobic	3	+ (1)	Reported as diazotroph/ scavenger	(Alarico et al. 2014; Gtari et al. 2012)
Xanthomonas	Both	3	+ (1)	Reported	(Biabani 2008)
Achromobacter	Aerobic	2	+ (1)	Yes	(Proctor and Wilson 1959)
Acinetobacter	Anaerobic	2	+ (1)	Yes	(Liba et al. 2006)
Agromyces	Both	2	-	Reported	(Sellstedt and Richau 2013)
Delftia	Aerobic	2	-	Yes	(Agafonova et al. 2017)
Hydrocarboniphaga	Aerobic	2	-		
Lysobacter	Aerobic	2	-	Yes	(Iwata et al. 2010)
Massilia	Both	2	-	Yes	(Bailey et al. 2014)
Paenibacillus	Anaerobic	2	+ (2)	Yes	(Liu et al. 2019)
Rhizobacter	Aerobic	2	-	Related to *Burkholderia*	
Roseateles	Anaerobic	2	+ (1)		
Williamsia	Both	2	-	Reported	(Trujillo et al. 2015)
Yonghaparkia	Aerobic	2	-		
Aeromonas	Anaerobic	1	-		
Arthrobacter	Anaerobic	1	-		
Comamonadaceae	Aerobic	1	-		
Cronobacter	Anaerobic	1	-		
Enterobacter	Anaerobic	1	-		
Herbaspirillum	Aerobic	1	-	Yes	

Klebsiella	Anaerobic	1	-	Yes	
Kocuria	Aerobic	1	-	Reported	(Abadi et al. 2021)
Leifsonia	Anaerobic	1	-		
Luteibacter	Aerobic	1	-		
Microbacteriaceae	Anaerobic	1	-		
Micromonospora	Aerobic	1	+ (1)	Plant beneficial	Trujillo et al
Mucilaginibacter	Anaerobic	1	-		
Nitrobacter	Aerobic	1	+ (1)		
Novosphingobium	Aerobic	1	-		
Parafrigoribacterium	Aerobic	1	-		
Phreatobacter	Anaerobic	1	-		
Sphingobium	Aerobic	1	-		
Sphingomonas	Aerobic	1	-		
Staphylococcus	Anaerobic	1	-		