University of New Hampshire
University of New Hampshire Scholars' Repository

Master's Theses and Capstones

Student Scholarship

Fall 2010

nifH diversity associated with Montastraea cavernosa identified using an optimized primer protocol

Nathan D. Olson
University of New Hampshire, Durham

Follow this and additional works at: https://scholars.unh.edu/thesis

Recommended Citation
Olson, Nathan D., "nifH diversity associated with Montastraea cavernosa identified using an optimized primer protocol" (2010). *Master's Theses and Capstones*. 591.
https://scholars.unh.edu/thesis/591

This Thesis is brought to you for free and open access by the Student Scholarship at University of New Hampshire Scholars' Repository. It has been accepted for inclusion in Master's Theses and Capstones by an authorized administrator of University of New Hampshire Scholars' Repository. For more information, please contact nicole.hentz@unh.edu.

nifH DIVERSITY ASSOCIATED WITH *MONTASTRAEA CAVERNOSA* IDENTIFIED USING AN OPTIMIZED PRIMER PROTOCOL.

BY

Nathan D. Olson
B.S., B.A., University of Hawai'i Hilo, 2008

THESIS

Submitted to the University of New Hampshire
in Partial Fulfillment of
the requirements for the Degree of

Master of Science
in
Microbiology

September, 2010

UMI Number: 1487006

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript
and there are missing pages, these will be noted. Also, if material had to be removed,
a note will indicate the deletion.

UMI 1487006
Copyright 2010 by ProQuest LLC.
All rights reserved. This edition of the work is protected against
unauthorized copying under Title 17, United States Code.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

This thesis has been examined and approved.

Thesis Director, Michael P. Lesser,
Research Professor of Molecular, Cellular and Biomedical Sciences

Elise R. Sullivan,
Clinical Assistant Professor of Molecular, Cellular and Biomedical Sciences

Stephen H. Jones,
Research Associate Professor of Natural Resources

8/2/10
Date

Acknowledgements

Research conducted for this thesis was supported by grants from the National Geographic Society, NOAA's National Institute of Undersea Science and Technology, the University of New Hampshire Marine Program, and the American Museum of Natural History. Additional support was provided by the University of New Hampshire through teaching assistantships and a summer graduate research fellowship. I would like to thank my committee for providing invaluable insight and support. Especially my thesis director, Dr. Michael Lesser, for providing me with the space, guidance, funding and support that allowed for me to play in the laboratory as well as the ocean the past two years.

Finally I would like to thank my family as they were always there to help me get through the hard times and celebrate the good times. Most importantly I would like to thank my wife Dr. Erica Seifert, it was with her strength and advice that I have been able to do things I never thought possible.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iii

LIST OF TABLES vi

LIST OF FIGURES vii

ABSTRACT viii

CHAPTER PAGE

INTRODUCTION 1

Biological Nitrogen Fixation 2

Evolution and Diversity of Nitrogen Fixing Bacteria 3

Advantages of Nitrogen Fixing Symbiosis 5

Coral Nitrogen Fixing Bacterial Symbiosis 6

Research Justifications 9

I. AN ANALYSIS OF UNIVERSAL *nifH* PRIMER MISMATCHES AND DEVELOPMENT OF AN ALTERNATIVE PROTOCOL FOR THE RECOVERY OF ENVIRONMENTAL SEQUENCES

Introduction 11

Methods 14

Results 20

Discussion 36

II. DIVERSITY OF NITROGEN FIXING BACTERIA ASSOCIATED WITH *MONTASTRAEA CAVERNOSA*

Introduction 41

Methods 44

Results 47

Discussion 55

III. CONCLUSION 59

REFERENCES 67

APENDIX 74

List of Tables

Chapter I

Table 1.1 Published universal *nifH* primer pairs

Table 1.2 Taxonomy of *nifH* sequence data set

Table 1.3 Physical characteristics of *nifH* primers

Chapter II

Table 2.1 BLAST results for representative OTU sequences

Table 2.2 Sample diversity estimates and coverage

List of Figures

Chapter I

Figure 1.1 Phylogeny of *nifH* sequence data set

Figure 1.2 Location of published *nifH* primers

Figure 1.3 Percent matches for universal *nifH* primer pairs

Figure 1.4 Percent matches for universal *nifH* primer pairs by cluster

Figure 1.5 Individual primer percent sequence match

Figure 1.6 Rarefaction curves for clone libraries

Figure 1.7 ChaoI richness estimates for clone libraries

Figure 1.8 Phylogeny of *nifH* sequence clone libraries

Chapter II

Figure 2.1 Comparison of individual OTUs by color and location

Figure 2.2 Comparison of individual OTUs for transition colonies

Figure 2.3 Whole *nifH* community analyses

ABSTRACT

nifH DIVERSITY ASSOCIATED WITH *MONTASTRAEA CAVERNOSA* IDENTIFIED USING AN OPTIMIZED PRIMER PROTOCOL.

by

Nathan D. Olson

University of New Hampshire, September, 2010

Thesis Advisor Dr. Michael P. Lesser

The diversity of nitrogen fixing bacteria in any system must be identified to in order to fully understand their ecological role. PCR is commonly used to investigate bacterial diversity. To capture the full diversity PCR primers must bind to and amplify all targeted DNA sequences. For this study I analyzed published universal *nifH* primers' ability to capture the full diversity of nitrogen fixing bacteria. Based on this work I developed a new protocol for capturing the full diversity of *nifH* sequences. Using this optimized protocol I investigated community differences in nitrogen fixing bacteria between orange and brown color morphs of the Caribbean coral *Montastraea cavernosa* among three geographic locations. Whole community analysis revealed no difference between morphs or location. However, specific groups of proteobacteria and cyanobacteria differed in abundance between the morphs, indicating specific bacterial groups are responsible for differences previously observed in fixation between color morphs.

INTRODUCTION

Nitrogen fixation, the ability to convert atmospheric nitrogen into ammonia, is a prokaryotic process restricted to members of the Bacterial and Archaeal domains. Biological nitrogen fixation (BNF) contributes significantly to the amount of new nitrogen available to a wide variety of terrestrial, aquatic, and marine organisms in nitrogen limited systems (Galloway et al. 1995; Falkowski 1997). Nutrient availability, especially nitrogen, influences the trophic biology and ecology of all organisms. In the marine environment, a carbon:nitrogen:phosphorus ratio of 106:16:1, known as the Redfield ratio, was described for open ocean planktonic primary producers. This ratio was long thought to reflect the nutrient requirements for phytoplankton growth (Redfield 1958), with nitrogen often cited as the limiting macronutrient. The role of nitrogen as a limiting nutrient over ecological timescales is a generally accepted, though a continually debated paradigm. Recent research has suggested that the Redfield ratio more accurately represents a global average of the planktonic community, rather than a specific requirement for the growth of phytoplankton (Klausmeier et al. 2004). Whereas the growth requirements for phyloplankton in a specific environment are governed by physical factors such as local oceanography (Li and Hansell 2008; Church et al. 2009), exogenous input from nutrient runoff, and aeolian deposition (Fanning 1989; Dong et al. 2000). Additionally, biological factors such as the microbial transformation of nitrogen by nitrogen fixation, nitrification, and denitrification have recently become more appreciated as processes influencing nutrient and specifically nitrogen stoichiometry (Arrigo 2005; Ward et al. 2007).

In the marine environment, nitrogen fixation was underestimated by early studies (Capone and Carpenter 1982), but more recently estimates for fixation are significantly higher with rates closer to that of terrestrial environments (90-130 Tg N yr^{-1}) (Galloway et al. 1995) or higher (Karl et al. 2002; Quigg et al. 2003). One reason for this underestimation was the low predicted contribution for the cosmopolitan genus of free-living marine nitrogen-fixing bacteria *Trichodesmium* spp. (Capone 1997). A majority of the observed gap in nitrogen budgets was filled with the discovery of large numbers of planktonic unicellular cyanobacteria (Zehr et al. 2001; Zehr et al. 2007). These discoveries, and the realization of the importance of marine sources of nitrogen in the global nitrogen budget, have highlighted the need for continuing research into the complex cycling of nitrogen in the marine environment (Zehr 2002). While nitrogen cycling is also influenced by anthropogenic impacts (Vitousek et al. 1997a) microbial transformations are fundamental to its cycling (Ward et al. 2007). Of these transformations nitrogen fixation plays a pivotal role in making the abundant but biologically unavailable form of nitrogen available to support the biosynthesis of amino acids and proteins in growing cells.

Biological Nitrogen Fixation

Nitrogen fixation is an energetically expensive process in which the two nitrogen atoms of a dinitrogen molecule are each reduced to NH_3. This process requires two enzymes that form the nitrogenase complex. The Fe protein is a homo-dimer encoded by the *nifH* gene, and has a 4Fe:4S core (Tripplett 2000). The Fe protein is reduced by the electron mediator, ferredoxin, and then reduces the second enzyme involved in the

fixation process, the Mo-Fe protein. The Mo-Fe protein is a four-subunit $\alpha_2\beta_2$ protein with $\alpha\beta$ dimers that are coded for by the *nifD* and *nifK* genes respectively, and contains a unique Fe/Mo-cofactor (Tripplett 2000). The Mo-Fe protein reduces the N_2 molecule and is then subsequently reduced by the Fe-protein, completing the cyclic process.

Evolution and Diversity of Nitrogen Fixing Bacteria

The process of nitrogen fixation is performed by a diverse group of prokaryotic microorganisms from both the Bacterial and Archaeal domains (Raymond et al. 2004). Nitrogenase is an evolutionarily ancient enzyme complex theorized to have evolved from an enzyme involved in the reduction of dinitrogen analogs (Raymond et al. 2004). The evolutionary history of nitrogen fixation and more specifically the genes responsible for fixation includes evidence of vertical and horizontal gene transfer events (Raymond et al. 2004). Vertical gene transfer is the duplication of a gene or an operon within a prokaryotic genome. This process is responsible for the presence of multiple copies of the *nif* operon in a single microbial genome, with as many as five copies of the operon found in a single genome (Wang et al. 1988). Horizontal gene transfer, the transfer of a gene or group of genes from one bacteria species to another, has caused evolutionarily unrelated species of bacteria to appear closely related based on *nif* phylogeny (Zehr et al. 2003). These instances of genes transfer play an intricate role in nitrogen fixing bacterial diversity. Additionally the ability of a microbial lineage to fix nitrogen is commonly lost on evolutionary times scales so closely related bacterial species do not always share the ability to fix nitrogen (Raymond et al. 2004). Therefore, general bacterial diversity studies are based 16S ribosomal gene sequences are not applicable to nitrogen fixing

bacterial diversity (Raymond et al. 2004). As a result genes which code for the subunits of the nitrogenase enzyme are targeted to evaluate diazotroph diversity (Zehr et al. 2003). Of these genes *nifH* is the most commonly targeted due to its conserved nature in comparison to *nifD* and *nifK*. However, some studies utilize *nifD* when a finer resolution of the nitrogen fixing bacterial community is desired (Prechtl et al. 2004). It is through the analysis of *nifH* gene diversity that researchers identified previously unknown, yet globally important groups of nitrogen fixing bacteria such as a ubiquitous group of pelagic unicellular cyanobacteria (Zehr et al. 2001).

While free-living nitrogen fixing bacteria are found in abundance in a number of ecosystems, a diverse group form symbioses with eukaryotic organisms (Kneip et al. 2007). The specificity of the symbiosis can vary from highly specific with one symbiont species associating with only a one host species, or a diverse symbiont with specific host, or specific symbiont and a diverse host. Highly specific symbioses are generally considered as more evolved associations, indicating greater dependence between host and symbiont (Douglas 1995). Only a few bacterial symbiont types are present in marine shipworms (Distel et al. 2002). In contrast to this specific symbiosis a diverse community of nitrogen-fixing bacteria associate with tropical marine sponges, with bacterial symbiont representatives from all major nitrogen-fixing bacterial taxonomic groups (Mohamed et al. 2008). For these symbioses it is clear that the shipworm symbiosis is more obligatory when compared to that of reef sponges.

Advantages of Nitrogen Fixing Symbiosis

The ability to establish a symbiotic association with nitrogen fixing bacterium allows eukaryotic organisms to escape ecological limitations in a variety of habitats. Organisms may be able to expand their spatial niches into low-nutrient environments, or their dietary niches to resources that have low combined nitrogen content, and can thus reduce competition or gain a competitive advantage. Organisms harboring symbiotic diazotrophs in mutualistic associations often exhibit highly evolved mechanisms for the efficient transfer of fixed nitrogen products. The benefits of this symbiosis to both host and symbiont have lead to the ubiquity of this type of association especially in nitrogen limited environments.

Nitrogen-fixing organisms are commonly found in terrestrial and marine environments with limited inorganic nitrogen resources (Zehr 2003). In the terrestrial environment lichens are arguably the best example; they are common not only in desert soil crusts (Eskew and Ting 1978; Belnap 2002), but also on rocks (Seneviratne and Indrasena 2006) and lava flows, where they help create soil and initiate the process of ecological succession (Crews et al. 2001; Kurina and Vitousek 2001). In all of these environments, both competitors and predators are scarce. Similarly, the associations between higher plants and rhizobia (e.g., in legumes) allow these plants to thrive in nitrogen-poor soils, a clear advantage over their competitors. Coral reefs are a marine ecosystem where it is advantageous to host nitrogen-fixing symbionts (O'Neil and Capone 2008). The high productivity of coral reefs in oligotrophic waters was first viewed as a paradox during Darwin's voyage on the *Beagle*. It is now understood that both plankton predation (Hamner 1995; Hamner et al. 2007) and nitrogen fixation,

coupled with efficient nutrient cycling, on and around reefs contribute significantly to the nitrogen requirements of reefs (Wiebe et al. 1975). Diazotrophs are common in the water column (Hewson et al. 2007), on the substrate (Larkum et al. 1988; Charpy et al. 2007), and in associations with invertebrates such as sponges (Mohamed et al. 2008), corals (Lesser et al. 2004) and possibly tunicates (Paerl 1984; Odintsov 1991). Reef building corals are the structural and trophic foundation of coral reef ecosystems; therefore our understanding of their ability to form symbiosis with nitrogen fixing bacteria is of primary importance when attempting to elucidate coral reef trophic ecology.

Coral Nitrogen-Fixing Bacterial Symbiosis

Reef building corals are composed of three primary compartments the skeleton, host tissue, and mucus. Each of these compartments are host to diverse symbiotic communities including members of all three domains of life as well as viruses (Knowlton and Rohwer 2003). The location of these symbionts in relation to the host reflects the evolved nature of the association. Endosymbiosis, where the symbiont lives within the host tissue, are more evolved associations than epibionts, those living on the tissue surface. Nitrogen fixing symbionts are associated with all three coral compartments, and found as both endo- and epibiont (Williams et al. 1987; Lesser et al. 2004; Chimetto et al. 2008).

The coral nitrogen-fixing bacterial symbiosis was first hypothesized when Williams et al. (1987) documented nitrogen fixation in the skeleton of *Acropora variabilis*. This research, however, did not investigate the diversity of symbionts or whether the host or symbiont gained any benefit from the association. Based on the results from light/dark

acetylene reduction assays it was hypothesized that the symbionts were photosynthetic cyanobacteria (Williams et al. 1987). Similarly, Odinstov (1987) observed nitrogenase activity in the hydrocoral *Millepora* where zooxanthellae, unicellular green algae and bacteria, were noted in the skeleton as well as the tissue. It was not clear from the studies conducted which member of the microbial consortium associated with the coral was fixing nitrogen and how similar this association might be to that of reef building corals (Odintsov et al. 1987). Further investigation into coral-diazotroph symbioses identified nitrogen-fixing bacteria of the class γ-proteobacteria associated with the skeleton of *Favia favus* (Shashar et al. 1994), which displayed higher rates of nitrogen fixation under illumination, and with exposure to glucose-enriched seawater. These results suggest that the nitrogen-fixing bacteria utilized glucose from either the host or algal co-symbiont as an energy source, implicating a potential benefit for the symbionts in this relationship.

More recently, terminal restriction fragment length polymorphisms (T-RFLP) 16S rRNA analysis identified a diverse and dynamic community of coral bacterial symbionts (Rohwer et al. 2002), including many possible nitrogen-fixing bacterial associates of reef building corals. However, it was not until 2004 that an endosymbiotic nitrogen-fixing bacterium was identified (Lesser et al. 2004). The cyanobacterial symbionts were hypothesized to transfer fixed nitrogen to the coral's algal symbionts, and fixed carbon in the form of glycerol to the cyanobacteria fueling nitrogen fixation (Lesser et al. 2007). Initial analysis of the bacterial diversity for this system indicates that the symbiosis is comprised of a single cyanobacterial species (Lesser et al. 2004). In comparison a much greater diversity of nitrogen-fixing bacteria, including cyanobacteria and proteobacteria, were associated with the Hawaiian corals *Montipora capitata* and *Montipora flabellata.*

The benefits to each partner in these symbioses as well as whether the bacteria are endo- or epibionts remains unknown; however, a correlation was found between Vibrionaceae *nifH* gene copy number and algal symbiont abundance for *M. capitata,* suggesting an interdependent relationship (Olson et al. 2009). A diverse community of bacteria were identified but a conserved phylogenetic cluster of bacteria in the Vibrionacea family were found only in association with *M. capitata,* and a less conserved cluster of γ-proteobacteria were associated with *M. flabellata*. This symbiont specificity may indicate coevolved, highly specific, symbiotic associations. Additionally nitrogen-fixing bacteria were found in association with the mucus of the Brazilian coral *Mussimilia hisipda* (Chimetto et al. 2008). Cultured *Vibrio* on nitrogen free media were identified as capable of nitrogen fixation through acetylene reduction assays, while the benefit to the host and symbiont in this association remains unknown (Chimetto et al. 2008; Tripp et al. 2010). The diversity of nitrogen-fixing bacteria identified with different coral species indicates the potential for a range of symbiotic associations between the coral host, its alga, and nitrogen-fixing symbionts which remains a largely unexplored area of research.

The state of knowledge of the coral nitrogen fixing bacterial symbiosis is plagued by the inconsistency of methods used to identify and quantify this association. Numerous studies used indirect methods for quantifying nitrogen fixation namely acetylene reduction. These studies revealed that a number of corals assayed exhibit signs of fixation. Additionally, all studies where methods that could identify bacteria potentially capable of fixing nitrogen were used found such bacteria. These studies commonly revealed a diverse community of bacteria capable of fixing nitrogen. However the inconsistency of methods used among the different studies prevents any direct

comparison between the communities identified. The main limiting factor in our understanding of this symbiosis is the disconnect between physiology and diversity as both have not been investigated for any one system.

Research Justifications

While the global diversity of free-living nitrogen fixing symbiosis is a continuing topic of numerous studies, the diversity of nitrogen fixing bacteria in symbiosis is also important in understanding the biogeochemical cycling of nitrogen. Our understanding of the role nitrogen fixing bacteria play in specific systems either on the ecosystem or organismal scale will play a pivotal role in advancing our understanding of nitrogen cycling. One such system in particular is the diversity of nitrogen fixing bacteria associated with reef building corals. As research has indicated corals form numerous symbiotic associations with nitrogen fixing bacteria making them an intriguing and important system for determining the underlining role of nitrogen fixing bacteria in nutrient cycling and symbiosis. The first step in understanding the role of a group of organisms including nitrogen-fixing bacteria is to identify the diversity of that group. While a number of studies have investigated the diversity of nitrogen fixing bacteria from a range of environments fewer studies have examined the efficacy of current approaches. The focus of my research discussed in the remainder of this thesis addresses two issues. First how well do currently utilized protocols recover the full diversity of nitrogen fixing bacteria within an environment? To address this question a combined *in silico* and *in vitro* approach was taken. Using the programming language PERL I identified the capacity of published universal *nifH* primers to bind to full length *nifH* sequences. Based on the results of this analysis I developed and tested an optimized protocol in the

laboratory by comparing the diversity captured to that of a commonly used protocol. Secondly, what is the diversity of nitrogen fixing bacteria associated with reef building corals, and more specifically, the Caribbean coral *Montastraea cavernosa*?

For the second question I utilized the optimized protocol to compare the community of bacteria potentially capable of fixing nitrogen associated with *Montastraea cavernosa*, between orange and brown color morphs as well as among three geographic locations.

CHAPTER I

AN ANALYSIS OF UNIVERSAL *nifH* PRIMER MISMATCHES AND DEVELOPMENT OF AN ALTERNATIVE PROTOCOL FOR THE RECOVERY OF ENVIRONMENTAL SEQUENCES

INTRODUCTION

Low concentrations of dissolved inorganic nitrogen (DIN) often limits primary productivity (Redfield 1958) while the most abundant form of nitrogen, dinitrogen (N_2), is generally biologically unavailable. Nitrogen fixation converts N_2 to biologically available ammonium. This process is carried out by organisms from both Archaeal and Bacterial domains (Young 1992; Raymond et al. 2004). Nitrogen fixing, or diazotrophic, bacteria play a fundamental role in ecosystem productivity, particularly ecosystems where low concentrations of DIN limit productivity (Zehr 2003). Because of their importance in providing new nitrogen, recovering the complete diversity of nitrogen fixers in microbial diversity studies is important in order to fully understand ecosystem productivity (Zehr 2003).

Investigations into microbial community ecology commonly utilize culture-independent approaches such as multi-template PCR reactions to identify the diversity of bacteria present in a community. For this approach, total DNA from an environmental sample is extracted and gene specific primers are used to amplify targeted gene sequences present within the sample. The efficacy of this approach is dependent upon the ability of the PCR primer to amplify all copies of a specific gene within a particular sample in an unbiased manner (Kanagawa 2003) .

This standard, however, is rarely realized so when employing PCR based techniques we should strive to remove or reduce as many inherent limitations as possible. The primary issue is PCR bias which is the preferential amplification of one template sequence over another (Polz 1994). There are two types of PCR bias; those that are inherent to the specific PCR protocol being used and those that are random (referred to as PCR drift). While research on how PCR bias effects microbial diversity estimates using the 16S gene has been done (Kanagawa 2003), there is little research on PCR bias for other genes commonly used in diversity studies. One such gene is *nifH*, which codes for the alpha subunit of the nitrogen fixing enzyme nitrogenase, and is commonly used to identify the diversity of nitrogen fixing bacteria within a community.

We can reduce random bias in *nifH* diversity studies by using replicate reactions where the products of multiple PCR reactions, amplified from the same template, are combined for use in downstream procedures such as cloning, T-RFLP, or DGGE (Polz and Cavanaugh 1998). While easily implemented, this approach is rarely used in *nifH* diversity studies (Steward et al. 2004). However, one important consideration is that methods for reducing the inherent bias employed in 16S diversity studies are not as easily applicable to *nifH* studies. This is primarily due to the fact that approaches used for 16S studies incorporate lowering the stringency of the PCR reaction by decreasing the annealing temperature in order to limit bias caused by mismatches between primer and template sequence (Polz and Cavanaugh 1998; Ishii and Fukui 2001). Due to differences in the requirements for diversity studies between *nifH* and 16S alternative approaches to reducing bias must be applied.

For *nifH* studies there are two main causes of inherent PCR bias that must be addressed. While *nifH* is the most conserved gene of the *nif* operon, degenerate PCR primers are required to successfully capture the full diversity of the diazotrophic community. The use of degenerate primers reduces the number of mismatches between template and primer sequences and in effect reduces bias. However, the use of degenerate primers is responsible for the first of the two main causes of PCR bias in *nifH* diversity studies. Degenerate primers are comprised of a number of primer sequences with varying GC content and correspondently varying annealing temperatures. PCR bias studies using 16S as a model gene have shown differences in GC content to cause bias (Polz and Cavanaugh 1998) but no similar bias has been detected when using the *nifH* primers nifH1 and nifH2 (Zehr and Capone 1996; Tan et al. 2003; Diallo et al. 2008). These studies used either southern blots, DGGE, or TRFLP to quantify the products of PCR reactions using templates of multiple known sequences of varying concentrations to evaluate potential biases. While no bias was observed for primers nifH1/nifH2, bias was observed using a nested primer set from Poly et al. (2001) as described by Diallo et al. (2008). Diallo et al. (2008) utilized DGGE to compare the abundance of PCR products for the three copies of *nifH* in the *Azotobacter vinelandii* genome. The results indicated a bias towards copies with mismatches in non-degenerate locations in the primer sequence.

A second cause of inherent PCR bias is primer mismatch. Primer mismatch, or the presence of a non-Chargaff base pairing between the primer and template sequence, significantly reduces the primer's binding affinity, thus promoting PCR bias (Sipos et al. 2007). Therefore primer mismatches should be minimized as a goal for the effective

recovery of all target genes from a mixed microbial community. While it is difficult to design universal *nifH* primers, one step in limiting inherent PCR bias is to reduce the presence of mismatches between primer and template sequences. The occurrence of *nifH* primer mismatch has been previously investigated and it was concluded that the primer set nifH1/nifH2 (Zehr and McReynolds 1989) is the most universal (Diallo et al. 2008).

Here we present our analysis of published *nifH* primers and their ability to match a taxonomically diverse set of *nifH* sequences, and therefore their ability to be considered universal. The work presented here is a logical follow-up to the work by Diallo et al. (2008) as it examines a larger set of primers and therefore provides a more comprehensive analysis of how *nifH* primer pair mismatches effect the recovery of phylogenetic diversity. A combined *in silico* and *in vitro* approach was employed to accomplish this objective by constructing and comparing data sets of published universal *nifH* primers and full-length *nifH* sequences. Additionally, we compared the clone libraries produced with two sets of nested primers using genomic DNA from a coral reef water column sample as a test template. A commonly used nested set of primers described by Zehr and Turner (2001) and a modified version of this nested set that was identified as having fewer mismatches with sequences in our analysis.

METHODS

nifH and primer sequence collection

A *nifH* primer data set was developed through an extensive search of the primary literature identifying 18 primer sets that were included in our analysis that are used to

Table 1.1. Published universal *nifH* primer pairs with the names of the forward and reverse primer sequences, amplicon length in base pairs (Amp.) based on average length of matches from 195 *nifH* sequence dataset.

Reference	Forward	Reverse	Amp.
Zehr and McReynolds	nifH1	nifH2	359
Ueda et al. 1995	19F	407R	384
Ohkuma et al. 1996	IGK	GEM	444
Ohkuma et al. 1996	IGK	YAA	462
Ohkuma et al. 1996	KAD	GEM	360
Ohkuma et al. 1996	KAD	YAA	378
Widmer et al. 1999	forA	Rev	461
Widmer et al. 1999	forB	Rev	368
Zani et al. 2000	nifH3	nifH4	471
Poly et al. 2001	PolF	PolR	360
Poly et al. 2001	PolF	AQE	336
Poly et al. 2001	Kadino	Emino	363
Poly et al. 2001	469	470	367
Marusina et al. 2001	F1	R6	443
Marusina et al. 2001	F2	R6	355
Mehta et al. 2003	M03F	M03R	387
Fedorov et al. 2008	nifH2f	nifH3r	213
Fedorov et al. 2008	F1	nifH3r	462
Sarita et al. 2008	nifHfor	nifHrev	470

target the whole diazotrophic community (Table 1.1). Primer degeneracy and Tm were calculated using FastPCR 6.0.112 beta with default settings. Sequences for the *nifH* data set were identified using four methods. The first two methods incorporated database searches of published genomes using the Entrez search query tool (www.ncbi.nlm.nih.gov/sites/entrez) and IMG (http://img.jgi.doe.gov/cgi-bin/pub/main.cgi). An Entrez query was used to search the GenBank genome database for gene names containing *nifH*. A similar search was performed using IMG, searching genomes in the database for *nifH*. The third method utilized an additional Entrez query searching the GenBank nucleotide database for *nifH* containing sequences between 500 and 5000 bp in length. Additional sequences used in a previous analysis of *nifH* primers (e.g. Diallo et al. 2008) and not identified by either of the three previous methods, were also included in the data set. This data set, totaling over 700 sequences, was further refined to include only full-length *nifH* sequences those including both start and stop codon sequences. Additionally, redundant sequences were identified and removed. Sequences were classified as redundant based on the criteria that identical sequences were only kept if they were a second copy of the gene within a genome or from a different species or strain. In addition to the *nifH* sequence, the taxonomic identity for each of the sequences was also retrieved from GenBank. This refined data set contained 195 full-length *nifH* sequences.

Phylogenetic analysis

The open reading frame for all 195 sequences in the *nifH* data set was translated

into amino acid sequences and aligned in MEGA4 (http://www.megasoftware.net/, Tamura et al. 2007) using clustalW algorithms. The resulting alignment was used to construct a 1000 bootstrap neighbor joining phylogenetic tree using the software MEGA4 with pair wise deletion of insertions and default settings.

***in silico* primer analysis**

Perl scripts (http://www.perl.org) were written to determine how well each of the primer sets matched sequences in the *nifH* dataset constructed as described above. Briefly, individual primers were scored on the number of mismatches between the primer and gene sequence. Individual primers received scores of 0, 1, or 2 for the number of mismatches. If there were more than two mismatches between the primer and sequence, no primer match was recorded. Additionally, for primer pairs where both primers had fewer than three mismatches, primer pair scores were determined as the sum of the mismatch score for the individual primers.

***in vitro* primer analysis**

Two nested PCR protocols were investigated for their ability to capture the full *nifH* diversity of an environmental sample. Triplicate clone libraries were produced for each protocol in order to identify instances of PCR drift. A water column sample was collected from Admiral Patch Reef, Key Largo Florida (25.0446°N, 80.3945°W) from a depth of 1–2 m. The sample was collected in a sterile 1L Nalgene bottle and transported on ice back to shore for initial processing. The water column microbial community was

collected on a 0.22 μm Millipore membrane filter (Millipore, Billerica, MA.), the filter was stored in DNA buffer (Seutin et al. 1991) and transported to the University of New Hampshire for further processing. Genomic DNA was extracted from the microbial community collected on the filter using a CTAB DNA extraction method (France and Kocher 1996). Briefly, the Millipore filter was incubated in 600 μl of CTAB extraction buffer (5mM cetyltrimethylammonium bromide (CTAB), 10 mM Tris, 2.5 mM ethylenediaminetetraacetic acid (EDTA), 1.4 M NaCl, 0.2% by volume β-mercaptoethanol) with 5 μl of 20 mg ml^{-1} proteinase K at 65°C for 3 h. The DNA was purified from the lysate using a standard chloroform purification step followed by ethanol precipitation and the DNA was eluted in molecular grade H_2O. The concentration of the purified DNA product was determined using a Thermo scientific NanoDrop 8000 spectrophotometer at 260 nm (Thermo Fisher Scientific Inc., Waltham, MA.)

The *nifH* gene was amplified from the extracted genomic DNA using one of two nested PCRs. The first employed the commonly utilized nested primer set nifH3/nifH4 in a first round reaction and nifH1/nifH2 in the second round (Zehr and Turner 2001). For the alternative nested reaction we substituted the primer IGK for nifH4 in the first round reaction, resulting in a nifH3/IGK first round primer set, which is the same primer set as YAA/IGK (Ohkuma and Kudo 1996). All PCR reactions were performed using the following reaction conditions: PCR reactions consisted 0.5 X TITANIUM Taq DNA Polymerase and 1 X buffer (Clontech, Mountain View, CA USA), 0.2 μM dNTP's, 0.4 μM of each forward and reverse primer, and 1 μg $μl^{-1}$ genomic DNA with a total volume of 25 μl. The cycling protocol consisted of a 3 min initial denaturation step at 95°C

followed by 30 cycles with a 95°C denaturation step, 53.5°C annealing step and a 72°C extension step all of which were 30 sec followed by a 7 min final extension step at 72°C. All PCR reactions were performed using an Eppendorf Mastercycler gradient (Eppendorf Hauppauge, NY).

PCR reaction products were gel electrophoresed in 1.5% agarose gel at 75V for 90 min. Bands of the appropriate size were excised and gel extracted using the QIAquick gel extraction kit (Qiagen Inc. Valencia, CA) according to the manufacturer's protocol except for the following modifications: the melted gel in QG buffer and isopropanol was allowed to incubate on the spin column overnight at 4°C and the elution buffer was prewarmed to 37°C prior to the final elution step. Gel extracted PCR products were prepared for ligation through the addition of 5' A's. For each reaction 0.5 X Titanium Taq and 1 X buffer were added to 0.5 µM dNTP's and gel extracted PCR product was added to bring the total volume to 10 µl and the reaction was incubated at 72°C for 30 min. The modified PCR product was then used in a ligation reaction of 50 ng pGEM T-Easy Vector (Promega Corp. Madison WI), 3 Weiss units T4 DNA ligase, 1 X T4 DNA ligase buffer, with PCR product added to bring the reaction volume to 10 ul, which was incubated overnight at 4°C. The ligated PCR product was cloned using the pGEM T-easy vector system (Promega Corp., Madison WI) according to the manufacturer's protocol. Transformed colonies were grown overnight in 1 ml LB broth, pelleted by centrifugation at 2,969 x g for 15 min and the resulting supernatant was removed. The bacterial pellets were sent to Functional Biosciences Inc. (Madison WI.) for plasmid prep and sequencing.

Sequence Analysis

Sequences were trimmed of vector sequence using the Ribosomal Database project website (http://rdp.cme.msu.edu/) and compared to the GenBank database using BLASTn to confirm identity. Sequences were translated *in silico* (http://searchlauncher.bcm.tmc.edu/seq-util/seq-util.html) and aligned using clustalW algorithms (MEGA 4.0). Based on the clustalW alignments, a distance matrix was created using the Prodist executable in the PHYLIP package (http://evolution.genetics.washington.edu/phylip.html). Rarefaction curves and ChaoI richness estimate calculations were performed using Mothur v.1.11.0 (http://schloss.micro.umass.edu/wiki/, Schloss et al. 2009). The chaoI richness estimate predicts the number of OTUs in a sample based on the number of number of singleton and doubleton OTUs in comparison to the total number of OTUs captured. The sequences obtained using the nested protocol were combined with the *nifH* dataset created in this study and the resulting sequence set was aligned using clustalW algorithms (MEGA 4.0). This alignment was then used to create a neighbor-joining tree (1000 bootstrap replicates) with MEGA4 in order to determine the taxonomic identity of the sequences obtained using the two nested PCR protocols. These sequences have been submitted to the GenBank (Accession Numbers HM601463-HM601542).

Results

***nifH* sequence dataset**

A data set of 195 full-length *nifH* sequences of known taxonomy from the National

Center for Biotechnology Information (NCBI) GenBank database was constructed. Based on previous *nifH* sequence analyses there are four primary phylogenetic clusters (Zehr 2003). We determined which cluster each of the sequences in our data set belongs to by aligning the amino acid translated sequences and that alignment was then used to build a phylogenetic tree for the cluster classification of each sequence (Fig. 1.1). The identity of the clusters was determined by comparing the taxonomic composition of the clusters to previously identified members of each of the clusters (Zehr 2003). A majority of the sequences (121 out of 195) in our data set are in Cluster I (Table 1.2) which is comprised of known nitrogen fixing aerobic and microaerophilic bacteria (Raymond et al. 2004). Six phyla of bacteria fall in Cluster I (Table 1.2) with most sequences belonging to Proteo- or Cyanobacterial phyla. Of the 195 sequences, 39 belonged to Cluster II, which is composed of obligate anaerobic bacteria primarily from the phyla Chlorobi, Firmicutes, and the class δ-proteobacteria (Table 1.2). The third cluster contains *nifH* sequences that encode for nitrogenases that use iron or vanadium in the reaction center, as alternatives to molybdum. This taxonomically diverse cluster contains bacteria from both Archaeal and Eubacterial domains (Raymond et al. 2004). Archaea dominate Cluster III represented by 18 sequences in our data set. Also within this cluster are *nifH* sequences from strains of the α-proteobacterium *Rhodopseudomans palustris,* which has multiple copies of *nifH* in its genome (Cantera et al. 2004). Finally, Cluster IV *nifH* sequences are dominated by Archaea, and the genus *Clostridium* within the Firmicutes (Table 1.2). Sequences in this cluster are not known to function in nitrogen fixation (Staples et al. 2007).

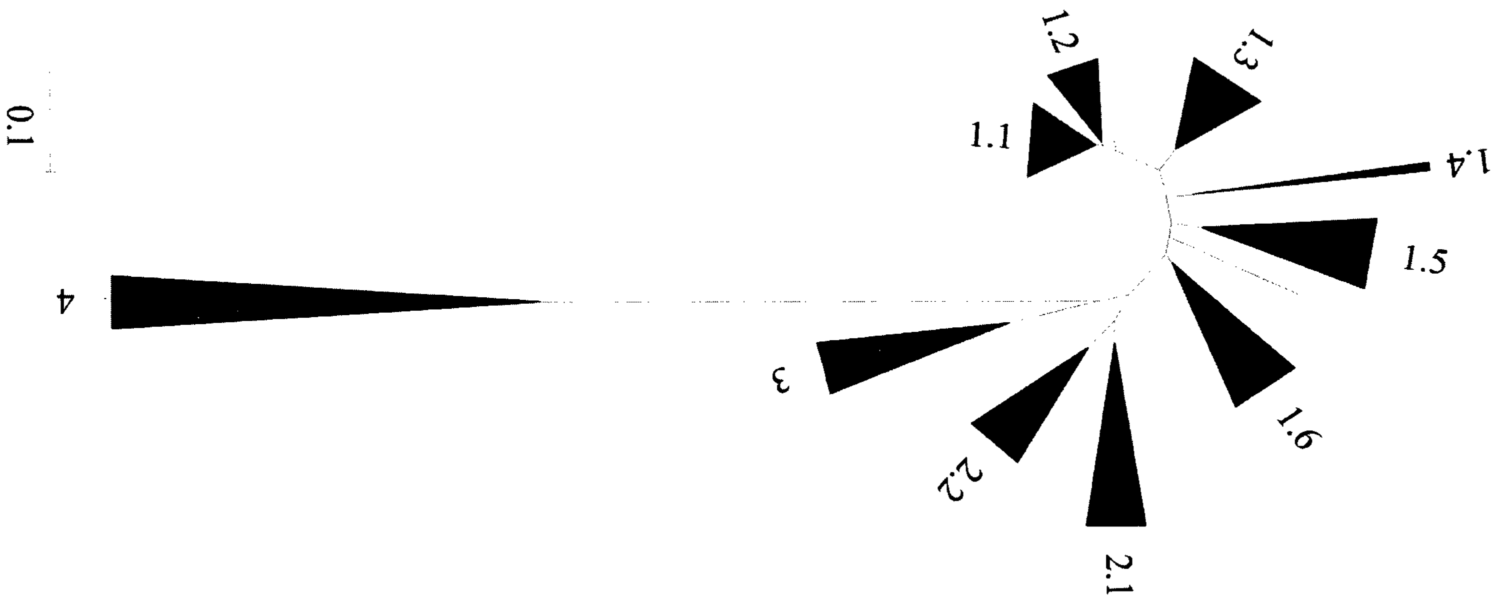


Figure 1.1. Phylogeny of *nifH* sequence data set used in analysis. Based on a neighbor joining (1000 bootstrap replicate) tree of 195 full-length amino acid *nifH* and *nifH* like sequences. Subgroups are labeled according to their cluster and numbered sequentially based on their order within that cluster. The base and height of the triangle for each subgroup or cluster is proportional to the number of and distance between the sequences respectively.

Table 1.2: Taxonomic breakdown of full-length *nifH* sequence data set used in analysis. Sequences were broken down into four *nifH* clusters as determined by phylogenetic analysis and phylum except for proteobacteria which was broken down further by class according to organisms known taxonomic identity.

Cluster	Phylum	Class	No. of Seq
I	Actinobacteria		5
	Cyanobacteria		16
	Firmicutes		9
	Nitrospirae		3
	Proteobacteria	α	56
		β	11
		δ	2
		ε	1
		γ	17
	Verrucomicrobia		1
			121
II	Chlorobi		11
	Chloroflexi		1
	Euryarchaeota		6
	Fibrobacteres		1
	Firmicutes		12
	Proteobacteria	δ	8
			39
III	Euryarchaeota		12
	Firmicutes		2
	Proteobacteria	α	4
			18
IV	Euryarchaeota		8
	Firmicutes		7
	Proteobacteria	α	3
			18
		Total	**195**

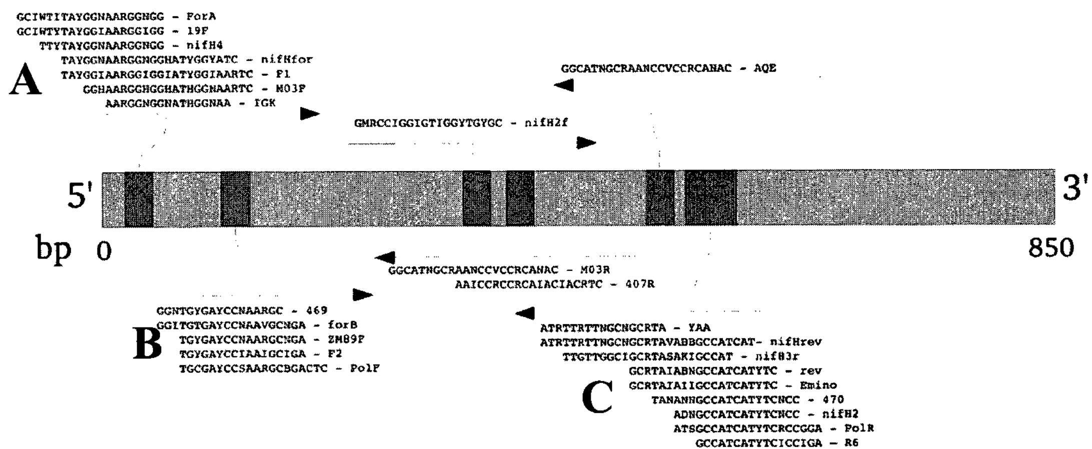


Figure 1.2: Location of published primers used in analysis based on alignment with *Azorhizobium caulinodans* ORS 571 AP009384. Primers are grouped together based or their position along the *nifH* gene with arrows representing the direction of the primers. Conserved regions of the gene where large groups of primers bind are labeled A-C. Dark regions along the gene are specific areas targeted by the individual groups of primers as indicated by the line and arrow.

***nifH* primer sets**

We analyzed a set of 18 universal primer pairs in this investigation. Most of these pairs target two of three conserved regions in the *nifH* gene (Fig 1.2), all of which correspond to the protein's ATP binding regions. The third region, C, is the most commonly targeted region of the three and is centered around three conserved amino acid residues (alanine, asparagine, and tyrosine) that interact with the adenosine ring of the ATP molecule (Burgess 2002). While these three regions are relatively conserved, the diversity of *nifH* genes in these regions still requires universal *nifH* primers to possess high levels of degeneracy (Table 1.3) with the F1 primer having the highest degeneracy, with over 4,000 individual primer sequences composing this primer (Marusina et al. 2001).

***In silico* primer analysis**

The comparison of the primer sets revealed a broad range of capability to recover the *nifH* diversity represented by our data set. The primer sets nifH1/nifH2 (Zehr and McReynolds 1989), 19F/407R (Ueda et al. 1995), all sets from Ohkuma and Kudo (1996), as well as Emino/Kadino and 469/470, both from Poly et al. (2001) were the most universal. These primers' sequences matched with over 80% of *nifH* sequences in our data set without any mismatches and over 90% when taking into consideration 1 to 2 mismatches (Fig. 1.3). Conversely, some of the primer pairs were not nearly as universal, but were less degenerate (e.g, nifHrev/nifHfor, Sarita et al. 2008). One of the most commonly used primer sets nifH3/nifH4 (Zehr and Turner 2001),

Table 1.3. Characteristics of primers analyzed. Primer names were kept the same as in the original publication except for when the primers were not named, in which case the primers were named according to the first letter of the last name of the first author, year published and direction (F= forward, R= reverse.) Nucleotide ambiguities correspond with standard code (R = A or G, K = G or T, S = G or C, Y = C or T, M = A or C, W = A or T, B = not A, H = not G, D = not C, V = not T, N = any nucleotide, I = inosine). Tm and degeneracy calculated using FastPCR 6.0.112 beta using default parameters.

Name	Sequence (5'-3')	Degeneracy	Length (bp)	Tm (°C)	Reference
nifH1	TGYGAYCCNAARGCNGA	128	17	54.4	Zehr and McRenolds 1989
nifH2	ADNGCCATCATYTCNCC	96	17	54.1	Zehr and McRenolds 1989
19F	GCIWTYTAYGGIAARGGIGG	1024	17	51.2	Ueda et al. 1995
407R	AAICCRCCRCAIACIACRTC	512	17	54.3	Ueda et al. 1995
IGK	AARGGNGGNATHGGNAA	384	17	51.2	Okuhuma et al. 1996
GEM	Same as ZM89R				Okuhuma et al. 1996
YAA	ATRTTRTTNGCNGCRTA	128	17	46.8	Okuhuma et al. 1996
KAD	Same as ZM89F				Okuhuma et al. 1996
forA	GCIWTITAYGGNAARGGNGG	2048	18	53.4	Widmer et al. 1999
forB	GCRTAIABNGCCATCATYTC	384	19	57.4	Widmer et al. 1999
rev	GGITGTGAYCCNAAVGCNGA	192	19	51.8	Widmer et al. 1999
nifH3	Same as YAA				Zani et al. 2000
nifH4	TTYTAYGGNAARGGNGG	128	17	49.9	Zani et al. 2000
PolR	ATSGCCATCATYTCRCCGGA	8	20	58.4	Poly et al. 2001
PolF	TGCGAYCCSAARGCBGACTC	24	20	60.7	Poly et al. 2001
AQE	GACGATGTAGATYTCCTG	2	18	62.5	Poly et al. 2001
Emino	GCRTAIAIIGCCATCATYTC	256	17	47.9	Poly et al. 2001
Kadino	Same as ZM89F with replacing I with N				Poly et al. 2001
469	GGNTGYGAYCCNAARGC	128	17	54.8	Poly et al. 2001
470	TANANNGCCATCATYTCNCC	512	20	53.1	Poly et al. 2001
F1	TAYGGIAARGGIGGIATYGGIAARTC	4096	22	56.1	Marusina et al. 2001
F2	Same as Kadino				Marusina et al. 2001
R6	GCCATCATYTCICCIGA	32	15	49.4	Marusina et al. 2001
M03F	GGHAARGGHGGHATHGGNAARTC	1296	23	57.7	Mehta et al. 2003
M03R	GGCATNGCRAANCCVCCRCANAC	768	23	61.6	Mehta et al. 2003
nifH2f	GMRCCIGGIGTIGGYTGYGC	1024	17	62.6	Fedorov et al. 2008
nifH3r	TTGTTGGCIGCRTASAKIGCCAT	128	21	59.4	Fedorov et al. 2008
nifHfor	TAYGGNAARGGNGGHATYGGYATC	768	24	59.1	Sarita et al. 2008
nifHrev	ATRTTRTTNGCNGCRTAVABBGCCATCAT	3456	29	60.8	Sarita et al. 2008

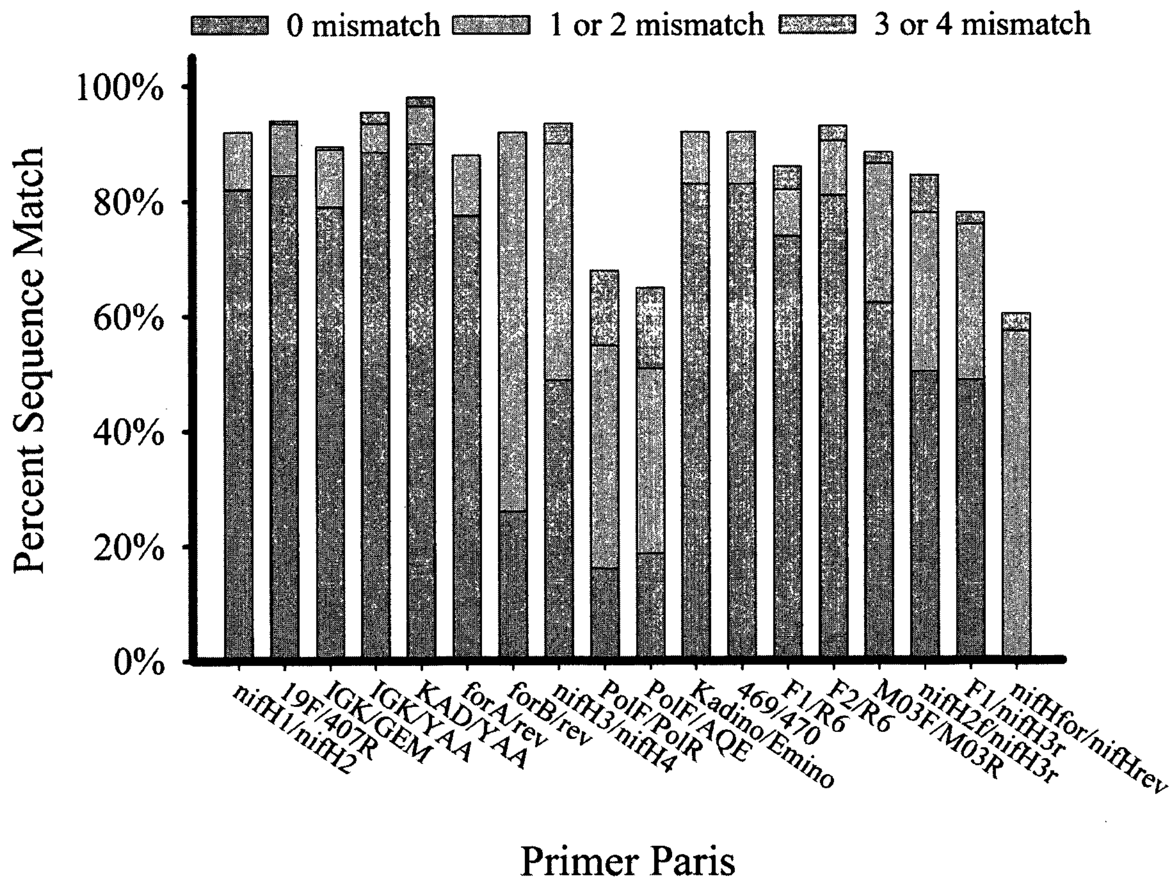


Figure 1.3 Percent matches for universal *nifH* primer pairs for all 195 sequences in the data set *nifH* phylogentic clusters. Each vertical bar is broken down by the percent of sequences that the primer pair matches the *nifH* sequence in the data base if taking into consideration no mismatches (blue), 1 or 2 mismatches (green), or 3 or 4 mismatches (red).

only matched with 50% of the sequences without any mismatches, and 90% when taking into consideration 1 or 2 mismatches. A primer pair's ability to capture the full diversity was related to when they were designed with the older primers being the most universal.

We were also interested in each primer set's ability to recover the diversity of the *nifH* clusters described above. All of the primer sets previously identified as universal recovered most of the diversity for each of the individual clusters. The exception was Cluster IV, where most matched fewer than 50% of the sequences, taking into consideration 1 or 2 mismatches (Fig. 1.4). There was also a decrease in the percentage of sequence matches from Cluster I to II and III for the older more universal primer pairs. Primer pair nifH3/nifH4 matched well with Cluster I but not as well with Clusters II and III (Fig. 1.4), and for Cluster II, they did not match with any of the sequences without mismatches. They matched with Cluster III sequences slightly better by matching with 30% of sequences without mismatches. The more universal primers not only matched up well with the sequence set overall but also did not match up well with non-functional *nifH* genes providing additional incentive for their use in studies targeting the diversity of bacteria that are actively fixing nitrogen.

***<u>In vitro</u>* <u>primer analysis</u>**

Our analysis of the nifH3/nifH4 primer set suggests it is not as universal as other primer sets investigated in this study (see above) and an analysis of the primer's ability to match sequences in our data set revealed that nifH3 is highly universal whereas nifH4

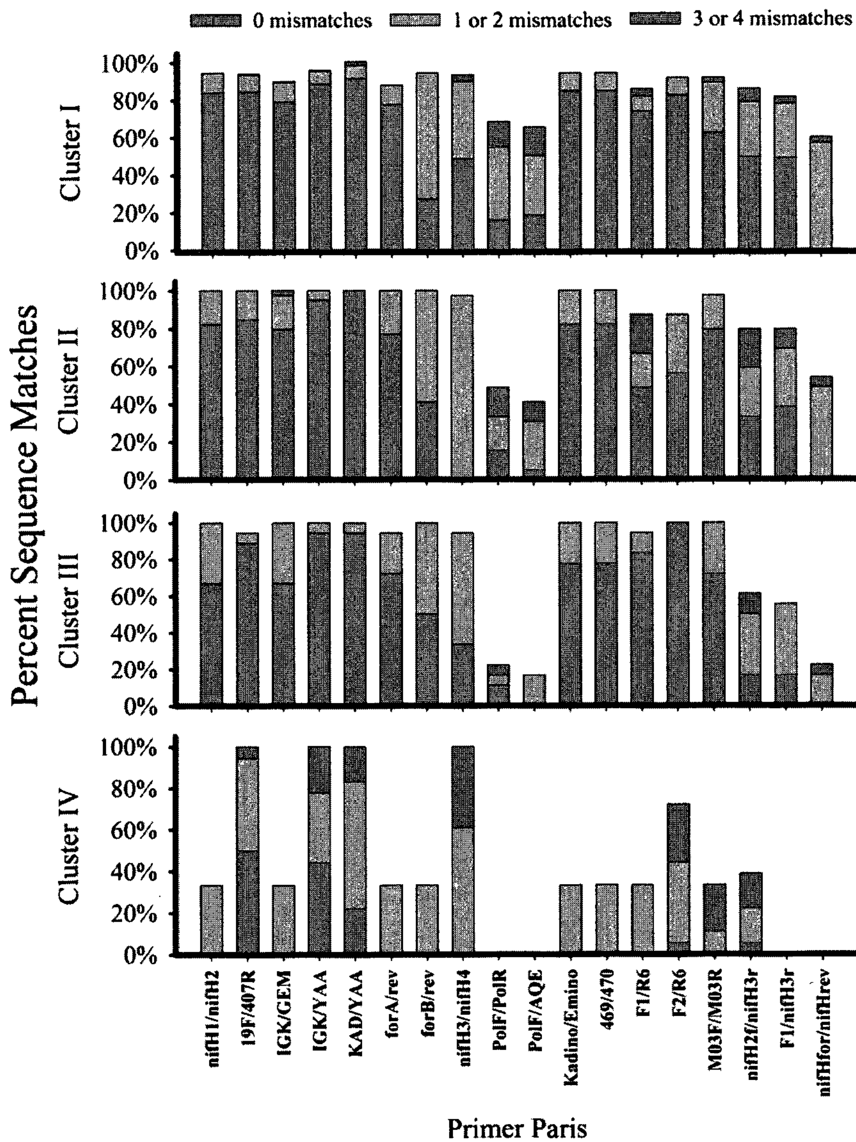


Figure 1.4: Percent matches for universal *nifH* primer pairs for the four *nifH* phylogentic clusters. Vertical bars are broken down by the percent of sequences that the primer pair matches the *nifH* sequence in the data base if taking into consideration no mismatches (blue), 1 or 2 mismatches (green), or 3 or 4 mismatches (red).

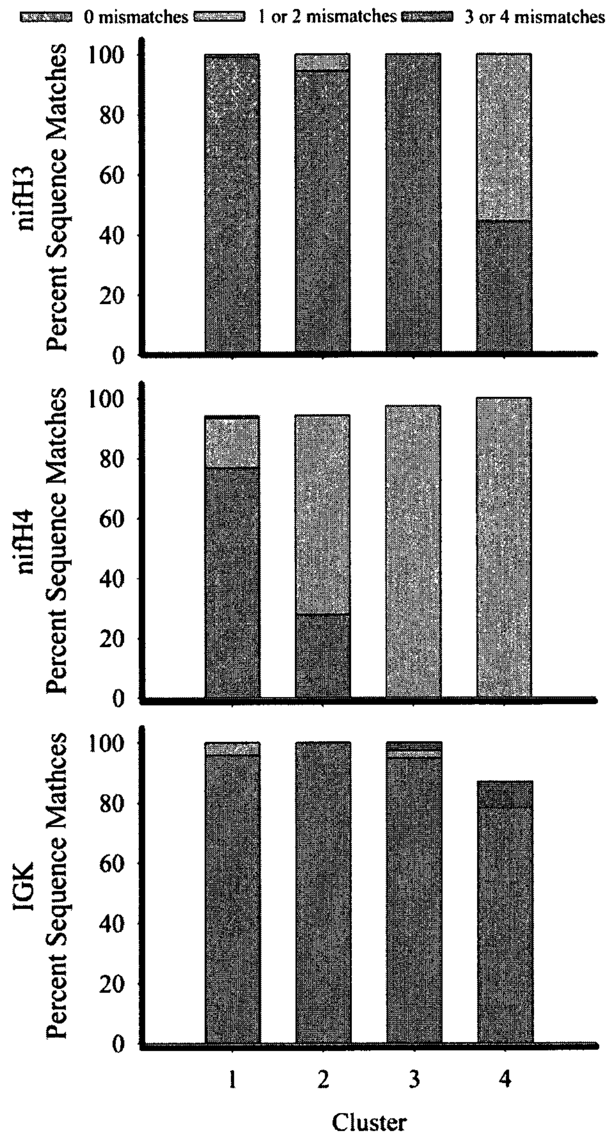


Figure 1.5: Percent sequence matches for individual primers broken down by *nifH* sequence cluster IGK (Okuhuma et al. 1996), nifH3 and nifH4 (Zani et al., 2001). Each vertical bar is separated by the proportion of the matched sequences requiring 0 (blue), 1 (green), or 2 (red) mismatches.

requires a greater number of mismatches to bind to sequences in Clusters I-III (Fig. 1.5). Alternatively, the IGK primer could replace nifH4 because it was identified as highly universal and targets the same conserved region (region A, Fig 1.2) of the *nifH* gene as nifH4. Therefore, we compared the most common nested protocol to an alternative protocol that only differed by utilizing the primer IGK instead of nifH4 in the first round reaction. This alternative protocol is the same primer set as YAA/IGK (Ohkuma and Kudo 1996). It is important to note that IGK is a more degenerate primer in comparison to nifH4 (Table 1.3). Our sequencing effort did not recover the total diversity of *nifH* genes amplified using the two protocols. However, rarefaction curves indicate that at the 95% OTU similarity cutoff both protocols are beginning to reach an asymptote (Fig 1.6) and chaoI richness estimates show no significant difference in diversity observed between the two protocols (Fig 1.7). While no difference in richness was observed between the two protocols only a fraction of the OTU's were shared between the two protocols, with 4 out of 23 and 3 out of 38 for OTU distance cutoffs at 95% and 99% respectively were shared.

Phylogenetic analysis of the sequences from the commonly used (N=39) and alternative (N=43) nested protocols revealed a diverse group of bacteria recovered belonging to both Clusters I and II (Fig.1. 8). The phylogenetic group identified as 1.4 (Fig 1.1) contained roughly equal numbers of sequences from both nested protocols (7 with nifH3/nifH4 and 6 with nifH3/IGK respectively) that were phylogentically most closely related to the ε-proteobacterium *Wolinella succinogenes* DS (Accession

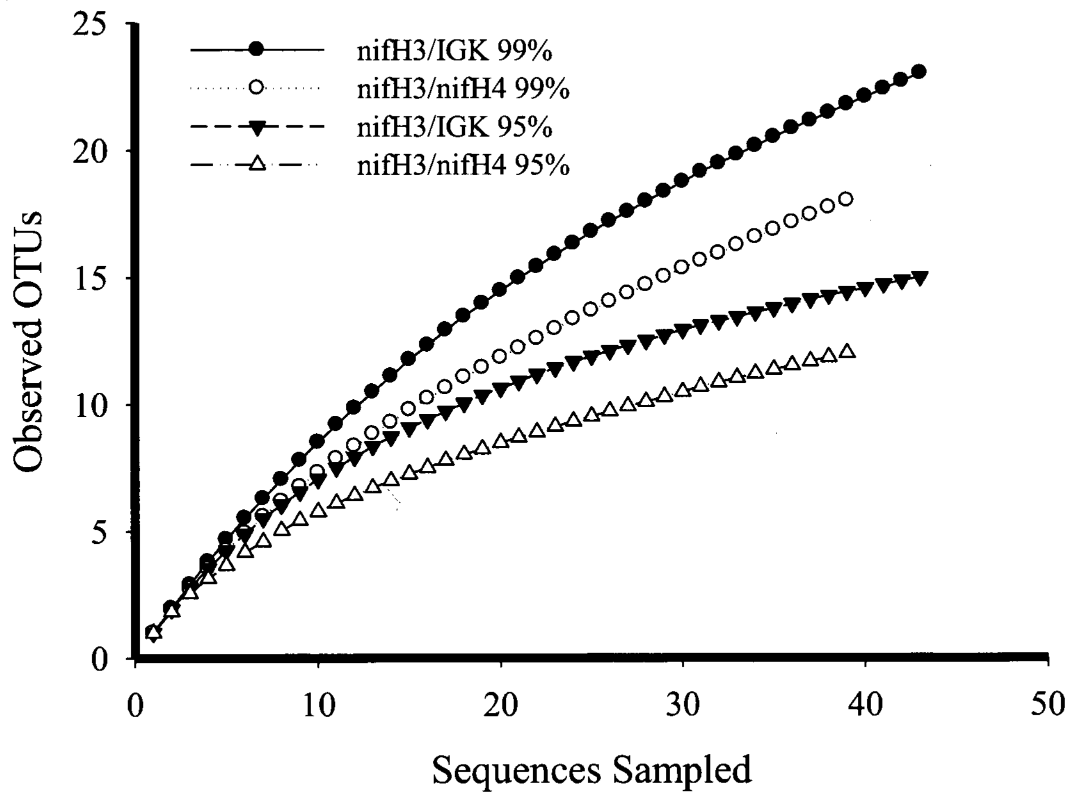


Figure 1.6: Rarefaction curve for pooled clone libraries for the nested protocol with nifH3/nifH4 and the alternative protocol with nifH3/IGK showing OTU cutoff of 99% (circles) and 95% sequence similarity (triangles).

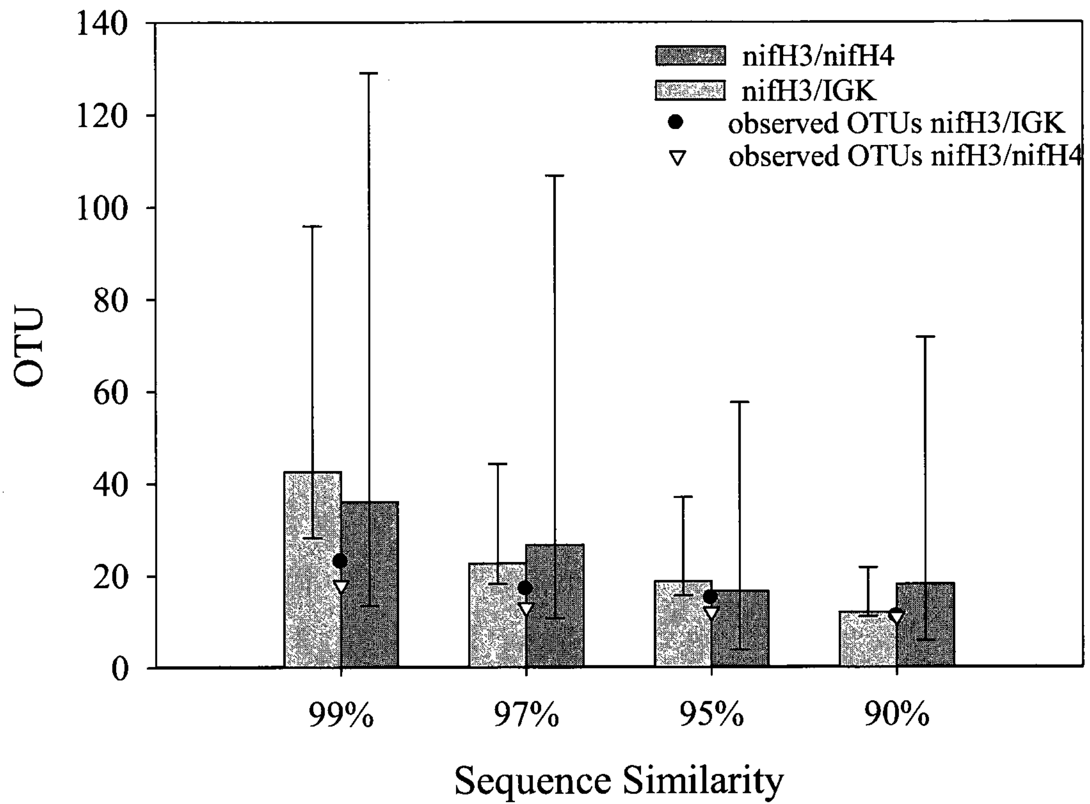


Figure 1.7: Chao1 estimates comparing the nested protocol with nifH3-nifH4 to the alternative protocol with nifH3-IGK, with observed OTUs for each cutoff. OTU cutoff's based on sequence similarity. High and low confidence intervals represented as error bars.

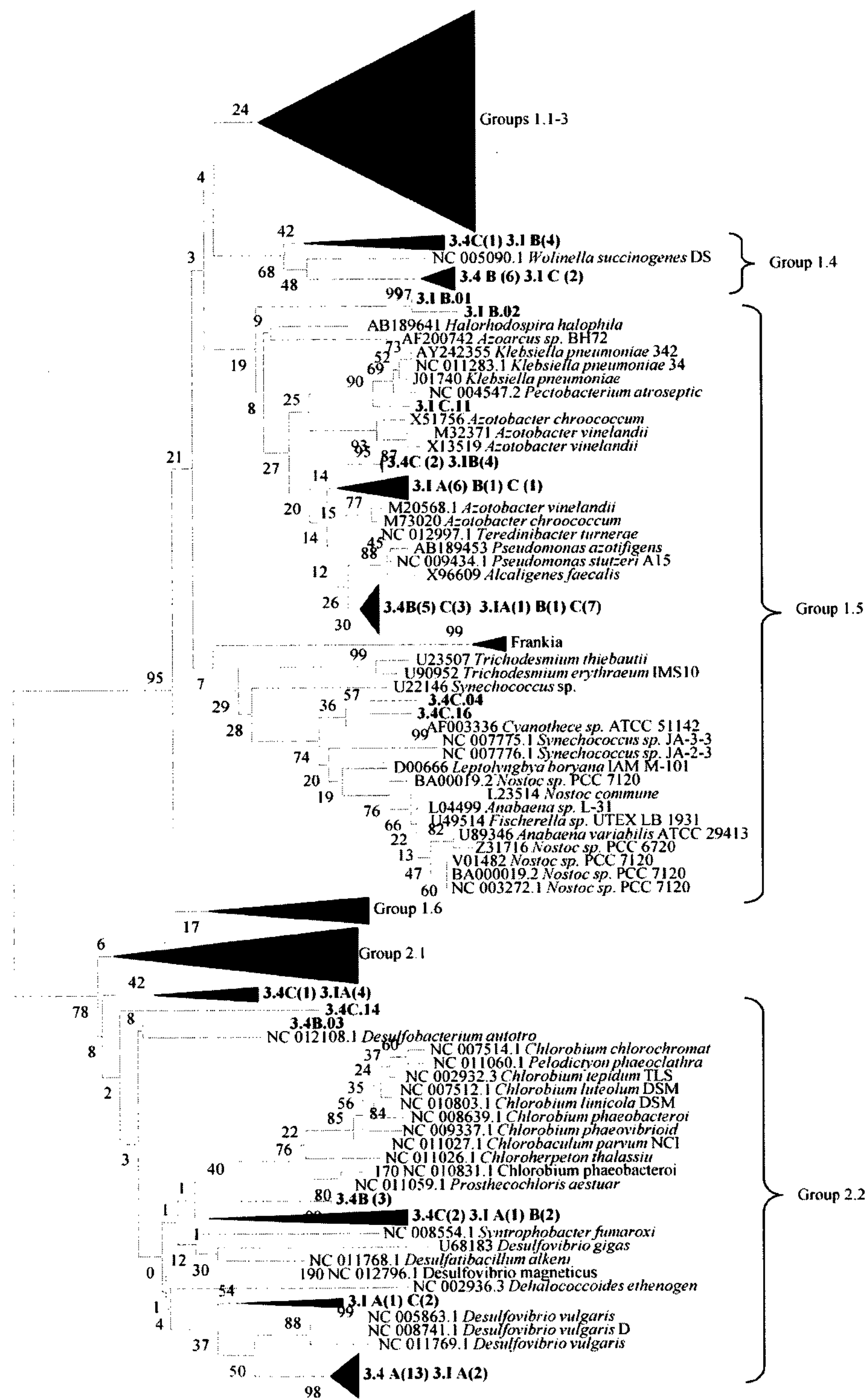

Groups 1.1-3
3.4C(1) 3.1 B(4)
NC 005090.1 Wolinella succinogenes DS
3.4 B (6) 3.1 C (2)
Group 1.4
3.1 B.01
3.1 B.02
AB189641 Halorhodospira halophila
AF200742 Azoarcus sp. BH72
AY242355 Klebsiella pneumoniae 342
NC 011283.1 Klebsiella pneumoniae 34
J01740 Klebsiella pneumoniae
NC 004547.2 Pectobacterium atroseptic
3.1 C.11
X51756 Azotobacter chroococcum
M32371 Azotobacter vinelandii
X13519 Azotobacter vinelandii
3.4C (2) 3.1B(4)
3.1 A(6) B(1) C (1)
M20568.1 Azotobacter vinelandii
M73020 Azotobacter chroococcum
NC 012997.1 Teredinibacter turnerae
AB189453 Pseudomonas azotifigens
NC 009434.1 Pseudomonas stutzeri A15
X96609 Alcaligenes faecalis
3.4B(5) C(3) 3.1A(1) B(1) C(7)
Group 1.5
Frankia
U23507 Trichodesmium thiebautii
U90952 Trichodesmium erythraeum IMS10
U22146 Synechococcus sp.
3.4C.04
3.4C.16
AF003336 Cyanothece sp. ATCC 51142
NC 007775.1 Synechococcus sp. JA-3-3
NC 007776.1 Synechococcus sp. JA-2-3
D00666 Leptolyngbya boryana IAM M-101
BA000019.2 Nostoc sp. PCC 7120
L23514 Nostoc commune
L04499 Anabaena sp. L-31
U49514 Fischerella sp. UTEX LB 1931
U89346 Anabaena variabilis ATCC 29413
Z31716 Nostoc sp. PCC 6720
V01482 Nostoc sp. PCC 7120
BA000019.2 Nostoc sp. PCC 7120
NC 003272.1 Nostoc sp. PCC 7120
Group 1.6
Group 2.1
3.4C(1) 3.1A(4)
3.4C.14
3.4B.03
NC 012108.1 Desulfobacterium autotro
NC 007514.1 Chlorobium chlorochromat
NC 011060.1 Pelodictyon phaeoclathra
NC 002932.3 Chlorobium tepidum TLS
NC 007512.1 Chlorobium luteolum DSM
NC 010803.1 Chlorobium limicola DSM
NC 008639.1 Chlorobium phaeobacteroi
NC 009337.1 Chlorobium phaeovibrioid
NC 011027.1 Chlorobaculum parvum NCI
NC 011026.1 Chloroherpeton thalassiu
NC 010831.1 Chlorobium phaeobacteroi
NC 011059.1 Prosthecochloris aestuar
3.4B (3)
3.4C(2) 3.1 A(1) B(2)
NC 008554.1 Syntrophobacter fumaroxi
U68183 Desulfovibrio gigas
NC 011768.1 Desulfatibacillum alkeni
NC 012796.1 Desulfovibrio magneticus
NC 002936.3 Dehalococcoides ethenogen
3.1 A(1) C(2)
NC 005863.1 Desulfovibrio vulgaris
NC 008741.1 Desulfovibrio vulgaris D
NC 011769.1 Desulfovibrio vulgaris
3.4 A(13) 3.1 A(2)
Group 2.2
0.05

Figure 1.8: Neighbor joining phylogentic tree (1000 bootstrap replicates) of Clusters I and II from *nifH* dataset and sequences from two nested protocols used in lab. Sequence groups are based on *nifH* sequence data set groupings (Fig 1.1). Sequences obtained using nested protocols are labeled according to the primer set used in the first round reaction either 3.4 for nifH3/nifH4 (N = 39) or 3.I for nifH3/IGK (N=43). Additionally each sequence is labeled according to which of the three replicate reactions the sequences are from (A, B, or C) with the number of sequences for each replicate indicated in parenthesis.

#NC005090.1). Based on our analysis both primer sets matched equally well with *W. succinogenes* DS (Appendix D). A second group of sequences were found closely related to γ-proteobacterium in group 1.6 (Fig 1.1) these sequences were primarily obtained using the nifH3/IGK primer set (21 versus 10 sequences). This group of γ-proteobacteria consistently had primer mismatches with the nifH3/nifH4 set but none with the nifH3/IGK primer set (Appendix E). There was a third group represented by 33 of the sequences obtained using the two protocols that grouped within Cluster II (Group 2.2, Fig 1.1), 21 were obtained with nifH3/nifH4 and 12 using nifH3/IGK. For all sequences in cluster II the nifH3/nifH4 primer set only matched when taking into consideration 1 or 2 mismatches where as the nifH3/IGK primer set did not require any mismatches except in the case of a single sequence (Appendix H). Of the 21 sequences obtained using the nifH3/nifH4 protocol 13 were from a single replicate reaction that was comprised of only a conserved group of sequences belonging to this group (Fig 1.8).

Discussion

Properly identifying the full diversity of nitrogen fixing bacteria within a community is fundamental to understanding biogeochemical cycling of nitrogen from the ecosystem to global scale (Zehr 2003; Ward et al. 2007). While PCR based approaches targeting *nifH* genes have significantly expanded our understanding of these microbial communities (Zehr et al. 2001), there are limitations to this approach that should be quantified as a basis for improving procedures to fully characterize the community of nitrogen fixing bacteria within a system. Here we present a new analysis of published *nifH* PCR primers and their frequency of mismatches to assess their ability to capture

taxonomically diverse *nifH* sequences from environmental samples.

We first characterized 18 published universal *nifH* primer pairs for their frequency of primer mismatch. In order to determine the frequency of mismatch we compared these primer sets to a data set of 195 *nifH* sequences that included representative sequences from all four previously identified *nifH* clusters. While sequences from all four major *nifH* clusters were represented, our data set was dominated by sequences belonging to Cluster I and was primarily composed of α-proteobacteria. This bias is due to the fact that the available data is primarily composed of cultured organisms with sequenced genomes, and does not necessarily reflect the composition of the numerous diazotrophs in the natural environment.

The first set of *nifH* primers was designed by Zehr and McReynolds (1989) and was designed using the limited number of available sequences at that time. Here, and in other studies, this primer set was identified as one of the most universal (Diallo et al. 2008). Since then, researchers have designed new *nifH* primer sets to limit primer mismatches which occur commonly when amplifying *nifH* using universal primers. The occurrence of primer mismatch is known to cause PCR bias with the potential to exclude rare, but important, members of the microbial community from diversity surveys (Sipos et al. 2007). While lowering the annealing temperature was found to reduce bias in 16S diversity studies involving primer mismatch (Ishii and Fukui 2001; Sipos et al. 2007), due to the high degeneracy of universal *nifH* primers this is not an option.

An investigation of the diversity of diazotrophic organisms within a rice paddy (Mårtensson et al. 2009) offers an example of primer mismatch causing bias in a *nifH*

diversity study. Using the nested primer sets PolF/PolR and PolF/AQE, a diverse group of proteobacteria was identified with both a clone library and DGGE. However, no cyanobacterial sequences were observed as previously identified in this system (Mårtensson et al. 2009). Subsequently, they were able to successfully obtain *nifH* sequences from a diverse group of cyanobacteria using a cyanobacterial specific primer set on their samples. In our analysis the nested primer sets used in their study had a higher occurrence of mismatches for cyanobacteria vs. proteobacteria (Appendix A-F) resulting in the lack of recovery of cyanobacterial sequences.

In the laboratory portion of our analysis there were three groups of sequences that we used as a basis for our analysis of the role of PCR mismatch in PCR bias (Fig. 1.8). The presence of primer mismatch for groups of sequences obtained using the two nested protocols described here can be assessed by examining the relative proportion of sequences in the clusters that were obtained using either of the two protocols. Sequences in Group 1.4 were equally represented in clone libraries using the two protocols. Based on our analysis, this was expected due to an equal number of predicted mismatches between primer and template sequences for the two primer sets. For the second set of sequences a greater proportion of the clone library for the nifH3/IGK primer set than the nifH3/nifH4 primer set was represented by sequences belonging to group 1.5. This disproportionate representation could reflects PCR bias caused by primer mismatch as nifH3/nifH4 primer set is predicted to require a greater number of mismatches to bind to template sequences in this group than the nifH3/IGK set. The third set of sequences in group 2.1, were also equally represented by sequences using both protocols. However,

our analysis predicted that the nifH3/nifH4 primer set required a greater number of mismatches to bind to the template sequences. These results contradict the hypothesized role of primer mismatch in PCR bias and PCR drift may provide an explanation in this case. Our sequence sets for the two protocols were based on three individual clone libraries from independent PCR reactions. One of the reactions using the nifH3/nifH4 protocol consisted solely of sequences from a conserved group of sequences. This reaction is an example of PCR drift where one template sequences is disproportionally represented in the pool of PCR products. So while, PCR mismatch plays a role in determining the inherent bias in PCR cases of PCR drift may also occur.

No significant difference in the estimated richness of *nifH* sequences were observed between the traditional and alternative protocol based on chao1 calculations (Fig 1.4). The two protocols yielded clone libraries composed primarily of OTUs unique to the protocols indicating taxonomic bias for the two primer sets. While the standard nested protocol was used to discover a ubiquitous and globally important pelagic unicellular nitrogen fixing cyanobacteria (Zehr et al. 2001), we suggest that this alternative protocol may lead to the discovery of additional globally important nitrogen fixing bacteria that were missed with the commonly used nested protocol due to the difference in the taxonomic bias between the two. As for all primer sets used to study the microbial composition of an environment one must take into consideration the limitations of the primers used and the targeted members of the community.

The results of this study will hopefully provide important information on the most common *nifH* primer sets that will assist in the determination of which PCR primers to

use in their diazotrophic diversity studies. Additionally, we hope that this work, along with previously published analyses, will serve as a starting point for research into the limitations and sources of bias in *nifH* diversity studies. While numerous studies have addressed bias in the optimization of their methods (Zehr and Capone 1996; Poly et al. 2001; Tan et al. 2003; Diallo et al. 2008) this study is the first to focus solely on the evaluation of *nifH* primers and issues concerning PCR bias.

CHAPTER II

DIVERSITY OF NITROGEN FIXING BACTERIA ASSOCIATED WITH *MONTASTRAEA CAVERNOSA*

Introduction

Nitrogen fixing bacteria form symbiotic associations with an array of hosts ranging from unicellular phytoplankton including diatoms and dinoflagellates to terrestrial plants such as legumes (Kneip et al. 2007). With the range of hosts comes a continuum of symbiotic relationships. Some of these symbiotic relationships are species specific while others are flexible and highly diverse. The specificity of the relationship provides insight into the obligatory nature of the symbiosis and in turn the evolution of the association (Kneip et al. 2007). To fully characterize a symbiotic relationship identifying the full diversity of host and symbiont is required.

Crossland and Barnes (1976) indentified nitrogen fixing bacteria associated with the tissue of the coral species *Acropora acuminate* and *Goniastrea australensis.* These results were based on acetylene reduction assays and did not reveal the identity of the organisms responsible for fixation. Based on the nature of fixation and the known microbial associates of the corals, cyanobacteria were accredited for the fixation. A second study utilizing acetylene reduction and DCMU (3,4-dichlorophenyl dimethylurea), an inhibitor of photosystem II, also identified nitrogen fixation in the reef building coral, *Acropora variabilis* (Williams et al. 1987). Again, the approach used in this study did not confirm the identity of the symbiont. However, the inhibition of

fixation when exposed to DCMU indicated that the bacteria responsible for fixation were phototrophic, most likely cyanobacterial. A later study also found photosynthetic dependent nitrogen fixation associated with a number of different coral species (Shashar et al. 1994). Additionally, they were able to culture a nitrogen-fixing bacterium from a coral imprint using nitrogen free media. The identity of the cultured symbiont was determined by southern hybridizations as closely related to the gamma proteobacterial species *Klebsiella pneumoniae*, indicating the presence of a heterotrophic nitrogen fixing symbiont. These first few studies focused more on characterizing the physiology of the symbiosis than the identity of the symbiont.

More recent studies have utilized culture-independent PCR based techniques which identified a diverse community of bacteria associated with reef building corals, including species of known nitrogen fixing bacteria. Most of these known nitrogen fixing bacteria belonged to the Phylum Cyanobacteria, coinciding with previous physiology studies (Rohwer et al. 2002). Metagenomic analysis of the coral *Porites astreoides* identified genes involved in nitrogen fixation primarily belonging to the Phylum Cyanobacteria (Wegley et al. 2007). However, application of the microarray geoChip 2.0 revealed a diverse community of heterotrophic and autotrophic nitrogen fixing bacteria associated with the coral species *Montastraea faveolata* (Kimes et al. 2010). Using *nifH* clone libraries one study focused primarily on the diversity of nitrogen fixing bacteria associated with Hawaiian corals *Montipora capitata* and *Montipora flabellate* (Olson et al. 2009). While species specific associations were tentatively identified a diverse group of both hetero- and autotrophic nitrogen fixing bacteria were found in association with

the coral hosts (Olson et al. 2009).

One system that has received recent attention in regards to nitrogen fixing symbionts is the Caribbean coral *Montastraea cavernosa* (Lesser et al. 2004; Lesser et al. 2007). Tissue sections were challenged with cyanobacterial FISH probes and immunoblots for nitrogenase and phycoerytherin were analyzed and identified an abundant group of nitrogen fixing cyanobacterial endosymbionts associated with the orange color morph of this species, but not the brown (Lesser et al. 2004). These symbionts were tentatively identified as belonging to the genus *Synecococcus* based on 16S rDNA sequence analysis (Lesser et al. 2004). Additionally, acetylene reduction analysis indicated that fixation occurs only with the orange color morph in a diel pattern with peaks in fixation at dawn and dusk (Lesser et al. 2007). These findings suggest the cyanobacteria are responsible for the difference acetylene reduction between the two color morphs. However, the diversity of nitrogen fixing bacteria associated with the two color morphs as a whole is unknown. It was the focus of this study to evaluate the diversity of nitrogen fixing bacteria between color morphs.

To investigate nitrogen fixing bacterial diversity *nifH* clone libraries were employed as it is a commonly used method for investigating the diversity of nitrogen fixing bacterial communities (see Chapter I, Introduction). This gene codes for a subunit of the nitrogenase enzyme, which is responsible for nitrogen fixation. However, the presence of a *nifH* sequence only indicates the presence of a bacterial species capable of nitrogen fixation and not the specific one that is actually fixing nitrogen.

Through analysis of the *nifH* sequences isolated from *Montastraea cavernosa* we identified the members of the nitrogen fixing bacterial community potentially capable of fixing nitrogen associated with the orange and brown color morph from three different geographic locations within the Caribbean. From this we were able to address questions regarding the diversity of this community as well as how the community differs between the color morphs. We were also interested in the variability in this association on a geographic scale, and whether any color morph specific associations were consistent among different geographic locations. To address this question we compared the diversity of *nifH* sequences between orange and brown color morphs and among three locations in the Caribbean; Florida Keys, Lee Stocking Island Bahamas, and Little Cayman Island.

Methods

Coral sample collection and initial processing

We collected all corals samples from a depth of 50 feet using SCUBA. Three orange and three brown *Montastraea cavernosa* colonies were sampled from Alligator Reef, Florida Keys, Florida, USA (24° 51'N, 80° 36'W), Rock Bottom Wall Reef Little Cayman (19°18' N, 81°16'W), and North Perry Reef Lee Stocking Island, Bahamas (23° 46' N, 76° 05' W). Additionally, two multi-colored colonies identified as in transition from brown to orange color morphs were sampled from North Perry Reef Lee Stocking Island, Bahamas. Samples were transported to shore where they were initially processed.

First loosely associated bacteria were removed from the coral using an airbrush to induce the production and removal of copious amounts of mucous (Lesser et al. 2007). Then individual polyps were removed from the fragment, preserved in DNA buffer (Seutin et al. 1991), and transported to the University of New Hampshire for further processing. For the multi-colored colonies brown, orange and transition polyps were separated and processed individually as previously described.

DNA extraction, PCR amplification, and cloning

Genomic DNA was extracted from the samples using a previously described CTAB protocol (Methods Chapter II,France and Kocher 1996). Genomic DNA concentrations were determined using a NanoDrop at 260 nm (Thermo Scientific, Waltham MA.). The following protocol was used to amplify the *nifH* gene from the extracted genomic DNA. A nested protocol was employed using the primer set YAA (5'-ATR TTR TTNGCN GCR TA-3') and IGK (5'-AAR GGN GGN ATH GGN AA-3') (Ohkuma and Kudo 1996) for the first round and nifH1 (5'-TGY GAY CCN AAR GCN GA-3') and nifH2 (5'-AND GCC ATC ATY TCN CC-3') (Zehr and McReynolds 1989) for the second round. The first round reaction consisted 0.5 X TITANIUM Taq DNA Polymerase and 1 X Buffer (Clontech, Mountain View, CA USA), 0.2 μM dNTP's, 0.4 μM forward and reverse primer, and 1 μg μl^{-1} genomic DNA with a total volume of 25 μl. The cycling protocol consisted of a 3 min initial denaturation step at 95°C followed by 10 cycles with a 95°C denaturation step, 53.5°C annealing step and a 72°C extension step all of which were 30 sec followed by a 7 min final extension step at 72°C. The

second round reaction consisted of 0.5 X TITANIUM Taq DNA Polymerase and 1 X Buffer (Clontech, Mountain View, CA USA), 0.2 μM dNTP's, 0.8 μM forward and reverse primer, and 2 μg μl^{-1} round one PCR product DNA with a total volume of 25 μl. The same cycle protocol as the first round was employed except for instead of a 10 cycle reaction 30 cycles were preformed. Second round PCR products were gel extracted and cloned as previously described in Methods Chapter II.

Sequence analysis

Sequences were initially trimmed of vector using the Ribosomal database pipeline (http://rdp.cme.msu.edu/). Sequences were compared to the GenBank database using BLASTn algorithm to confirm identity. The open reading frames of the *nifH* sequences from all clones were translated into their amino acid sequences using the bacterial codon table. Sequences from the clone libraries were aligned in MEGA4 using the clustalW algorithm. A neighbor joining phylogenetic tree with 1000 bootstrap replicates was produced using the aligned *nifH* sequences. A distance matrix was calculated using the PHYLIP software package (http://evolution.genetics.washington.edu/phylip.html). The distance matrix was used to assign operational taxonomic units (OTUs), from which richness estimates and diversity indices were calculated using mothur v 1.11.0 (www.mothur.org). ChaoI estimates were used to predict sample richness, this calculation is based on the proportion of OTUs captured represented by one or two sequences to the total number of OTUs. Sample coverage was determined by dividing the number of captured OTUs by the ChaoI estimate for the sample. The Shannon

diversity index was used to determine the evenness of a population with a value of 0 indicating evenly represent members of the community and evenness decreasing with increasing values. The OTU abundance's for pooled samples were calculated and statistically compared using metastats (http://metastats.cbcb.umd.edu/). The phylogenetic tree produced was used for whole community analysis completed with UniFrac (http://bmf2.colorado.edu/unifrac/index.psp).

All *nifH* sequences obtained in this study were submitted to genBank (accession number HM999081-HM999629).

Results

A total of 549 *nifH* sequences were obtained from the 24 samples analyzed. The number of clones sequenced per sample ranged from 10 to 33. A diverse community of nitrogen fixing bacteria dominated by proteobacteria was found in association with *M. cavernosa* (Table 2.1). These results were consistent with previous studies that investigated the diversity of nitrogen fixing bacteria associated with reef building corals (Olson et al. 2009; Kimes et al. 2010). For each of the individual samples the number of observed operational taxonomic units (OTUs), Chao1 richness estimate, Shannon diversity index and percent coverage were determined (Table 2.2). The richness estimate values varied greatly among samples with values ranging from 2 to 32 OTUs based on a 95% sequence similarity cutoff for OTU clustering. The depth of the sequencing effort is reflected in the percent coverage with values ranging from 44% to 100%. Samples with 100% sequencing coverage all had estimated richness values of 4 or less, indicating

Table 2.1: BLAST results for representative sequences of OTUs including more than one sequence from more than one sample. N indicates the number of sequences in the data set belonging to the OTU. BLAST results based on tBlastn algorithm, with the accession number, description, and score for the best matched sequence along with the Division of the closest match of known taxonomy.

OTU	N	Best Matched Organism	Acc #	Score	Division
1	29	Uncultured bacterium clone Olson_A2-8-E0	EU693413	8e-41	Cyanobacteria
2	25	Uncultured bacterium clone TM2cl3	GQ464078	4e-41	α-proteobacteria
3	61	Uncultured bacterium clone OilUp-10319	DQ078039	2e-41	γ-proteobacteria
4	12	Uncultured nitrogen-fixing bacterium clone C229	DQ098258	1e-41	γ-proteobacteria
5	7	*Lyngbya majuscula*	AY115593	2e-39	Cyanobacteria
6	11	Uncultured bacterium clone Olson_C2-9-D08	EU693435	3e-40	α-proteobacteria
7	3	Uncultured bacterium clone IS1H10	EU594054	1e-41	Verrucomicrobia
8	12	Uncultured bacterium clone Mf20-A06-E01	EU693395	3e-41	γ-proteobacteria
9	15	Uncultured nitrogen-fixing bacterium A37	AF099781	4e-39	α-proteobacteria
11	4	Uncultured bacterium clone Mf21-A02-F05	EU693398	3e-39	γ-proteobacteria
12	10	Uncultured bacterium clone Mf21-A02-F05	EU693398	6e-41	γ-proteobacteria
13	2	Uncultured bacterium clone ML4G06	EU594212	5e-40	Cyanobacteria
14	11	Uncultured nitrogen-fixing bacterium isolate DGGE gel band H18	GU367893	4e-40	γ-proteobacteria
15	4	Uncultured nitrogen-fixing bacterium clone B215	DQ098223	9e-42	δ-proteobacteria
17	12	Uncultured bacterium clone RT_MLH04	EU594243	1e-40	Cyanobacteria
18	24	Uncultured microorganism clone A28	GQ241892	7e-37	Firmicute
19	114	*Bradyrhizobium jicamae* strain LMG 24556	HM047127	1e-40	α-proteobacteria
20	20	*Chloroherpeton thalassium* ATCC 35110	CP001100	1e-41	Chlorobi
21	9	Uncultured bacterium clone IS1H10	EU594054	5e-42	α-proteobacteria
22	2	Uncultured soil bacterium clone hrb20	GU111795	3e-41	α-proteobacteria
23	12	*Bradyrhizobium denitrificans* strain LMG 8443	HM047125	1e-40	α-proteobacteria
24	2	Uncultured microorganism clone H01_DNA_E3	EF568532	6e-41	δ-proteobacteria
26	7	Uncultured microorganism clone 20053A08	EF174696	1e-36	γ-proteobacteria
27	7	Uncultured bacterium clone 3H50	DQ177013	2e-37	δ-proteobacteria
28	19	*Vibrio cincinnatiensis*	AF134809	2e-40	γ-proteobacteria
30	7	Uncultured bacterium clone 3-62	FJ807378.1	2e-25	γ-proteobacteria
31	4	*Mesorhizobium tianshanense*	FM203332	2e-29	α-proteobacteria
34	3	Uncultured bacterium clone Mf21-C10-C09	EU693444	4e-43	Chlorobi
35	37	Uncultured bacterium clone 3-62	FJ807378	6e-28	γ-proteobacteria
37	6	Uncultured bacterium clone A5-H-19	FJ686505	8e-41	γ-proteobacteria
38	3	Uncultured bacterium clone 3H19	DQ177037	1e-40	Actinobacteria
45	6	Uncultured nitrogen-fixing bacterium isolate DGGE gel band SY13	EF196648	2e-40	δ-proteobacteria
47	3	Uncultured bacterium clone II-0.77	GU192801	3e-41	γ-proteobacteria
50	3	*Vibrio natriegens*	AF082989	6e-41	γ-proteobacteria
52	9	Uncultured bacterium clone 3-62	FJ807378	3e-41	γ-proteobacteria

Table 2.2: Diversity estimates for all samples based on OTUs with 0.05 sequence distance cutoff. Number of clones sequenced for each sample (N), number of observed, Chao1 richness estimates, Shannon diversity index, as well as sample percent coverage. Percent coverage based on the number of observed OTUs/ChaoI richness estimate.

Sample	N	OTUs	ChaoI	Shannon	% Coverage
Florida Keys					
OF1	33	11	25.0	1.6	44.0
OF2	10	4	4.0	1.2	100.0
OF3	23	4	4.0	1.2	100.0
BF1	15	2	2.0	0.0	100.0
BF2	38	17	32.0	2.4	53.1
BF3	23	9	11.0	1.8	81.8
Little Cayman					
OLC1	23	11	21.5	2.1	51.2
OLC2	17	8	18.0	1.8	44.4
OLC3	16	9	14.0	2.0	64.3
BLC1	17	9	19.0	2.0	47.4
BLC2	18	5	8.0	1.1	62.5
BLC3	26	4	4.0	0.9	100.0
Lee Stocking Island					
OLS1	19	3	3.0	0.5	100.0
OLS2	26	8	13.0	1.5	61.5
OLS3	26	7	17.0	1.1	41.2
BLS1	26	5	6.0	1.0	83.3
BLS2	27	3	3.0	0.6	100.0
BLS3	28	4	5.0	0.6	80.0
Transition					
B1	11	6	7.5	1.6	80.0
B2	19	9	10.5	1.9	85.7
T1	20	13	18.6	2.0	69.9
T2	20	6	7.5	1.4	80.0
O1	21	10	17.5	2.0	57.1
O2	21	7	10.0	1.5	70.0

communities with low richness. In addition to percent coverage the Shannon index of diversity was also calculated, with higher indices indicating greater evenness for the samples. Evenness was greater for samples having higher observed and estimated OTUs. Conversely, samples with lower estimates, and ultimately higher coverage tended to have low diversity indices, reflecting unevenness in the samples. Some examples of this unevenness are samples BF1, BLS2 and 3 (Table 2.2).

Based on the statistical analysis of the mean percent abundance between the color morphs 7 OTUs were found statistically different between the two color morphs when all samples for the two color morphs were pooled. Of these seven, five had a greater mean abundance in the orange versus brown color morphs (Fig 2.1). These include OTUs 17, 35, 37, 45, 52 which all belong to the bacterial class γ-proteobacteria except for OTU 17 and 45 which are in the Phylum Cyanobacteria and class δ-proteobacteria respectively (Table 2.1). The two OTUs that were more abundant for the brown color morph were 1 and 6 which are Cyanobacteria and α-proteobacteria respectively both of which matched up mostly closely with *nifH* sequences obtained from the coral species *Montipora flabellata*. OTU's 35 and 37 were statistically different when comparing all orange versus brown color morphs. When comparing the communities between the two color morphs at specific locations OTU 35was only significant for Lee Stocking Island, and OTU 37 for Little Cayman. OTU 17, however, was found in a higher abundance for the orange color morph for all three locations and when present the other OTUs (35, 37, 45, and 52) were also in a higher abundance (Fig 2.1). Sequences in OTU 17 were most closely related to sequences obtained from coral reef sponges, and most closely related to

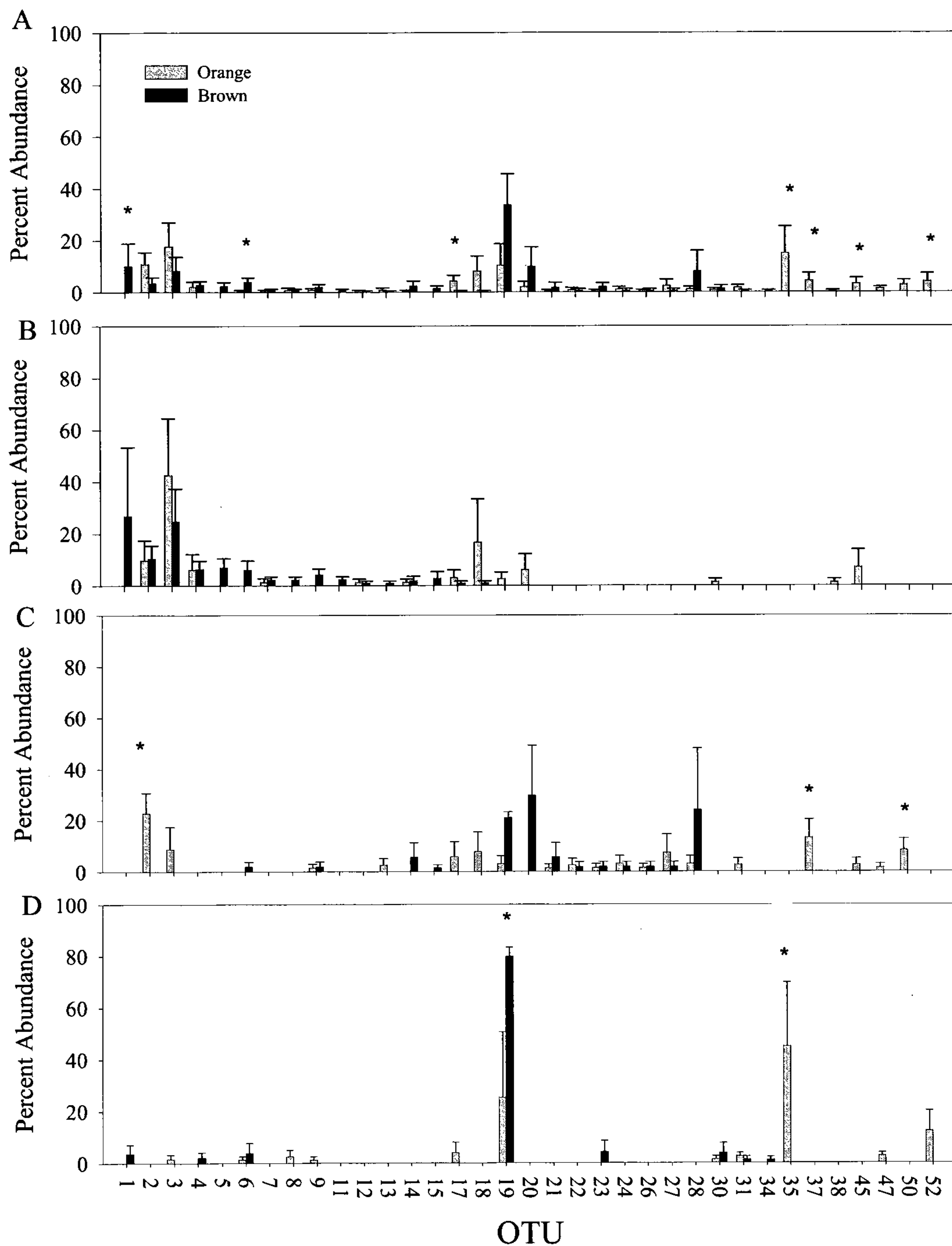


Figure 2.1: Color and location comparison for the percent abundance for individual OTUs. Comparison of orange and brown samples pooled for all locations (A), Florida Keys (B), Little Cayman (C), Lee Stocking Island (D). Asterisks indicate OTUs where the percent abundance is statistically different between the pooled samples.

the cyanobacterial genus *Cyanothece*. No individual OTU's were statistically different between color morphs for the Florida Keys samples where as 3 and 2 were statistically different for Little Cayman and Lee Stocking Island samples respectively. OTUs 2, 37, and 50 had a greater abundance for orange versus brown colonies for Little Cayman samples. With OTU 2 and 50 belonging to the class α- and γ- proteobacteria respectively (Table 2.1). For Less Stocking Island OTU 19, an α-proteobacteria, was more abundant in brown colonies and OTU 35 was more abundant in orange colonies.

When comparing individual polyps from the transition colonies three OTU's were found statistically different 14, 18, and 21 (Fig 2.2). OTU 14, γ-proteobacteria, was only present in orange and transition polyps and its abundance was statistically different between the orange and brown polyps. OTU 18 is in the Phylum Firmicutes and is more abundant in the brown polyps than transition and orange polyps but the difference was only statistically different between the orange and brown polyps (Fig 2.2). OTU 21, comprised of bacteria from the class α-proteobacteria, was only present in transition polyps and brown polyps, with the abundance of both being statistically different from the orange polyps.

We were also interested in potential differences in the nitrogen fixing bacterial communities as a whole. Using UniFrac we were able to compare the community structures among the different samples. When looking at the samples individually based on principal component analysis the samples did not group in a discernable pattern either based on color or location (A-C, Fig 2.3). However, when samples are pooled by

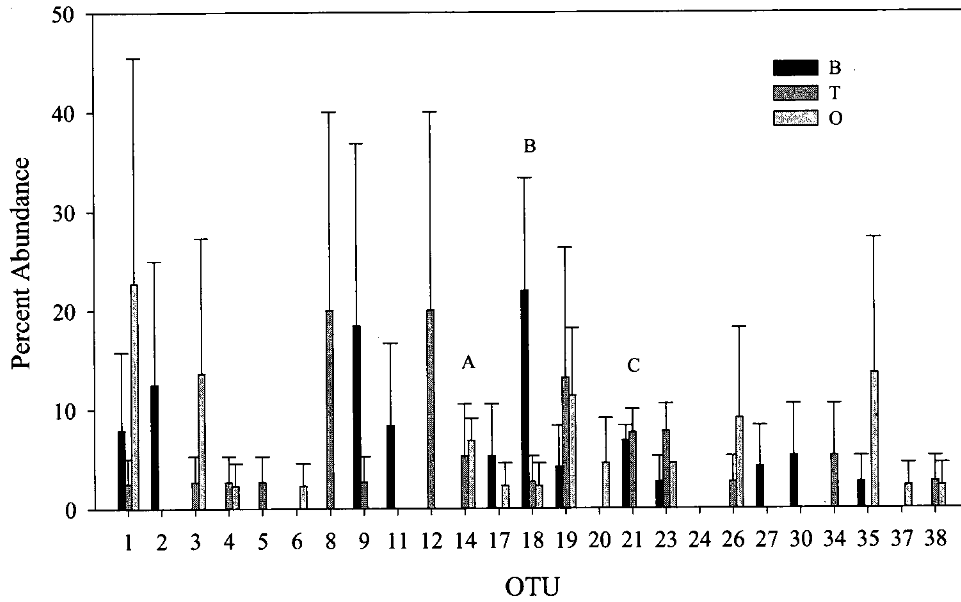


Figure 2.2: Comparison of percent abundance for the pooled replicate samples for transition colonies; orange (O), transition (T), and brown (B) polyps. Statistical differences between; B and O, B and T, and O and both B and T indicated as A, B, C respectively.

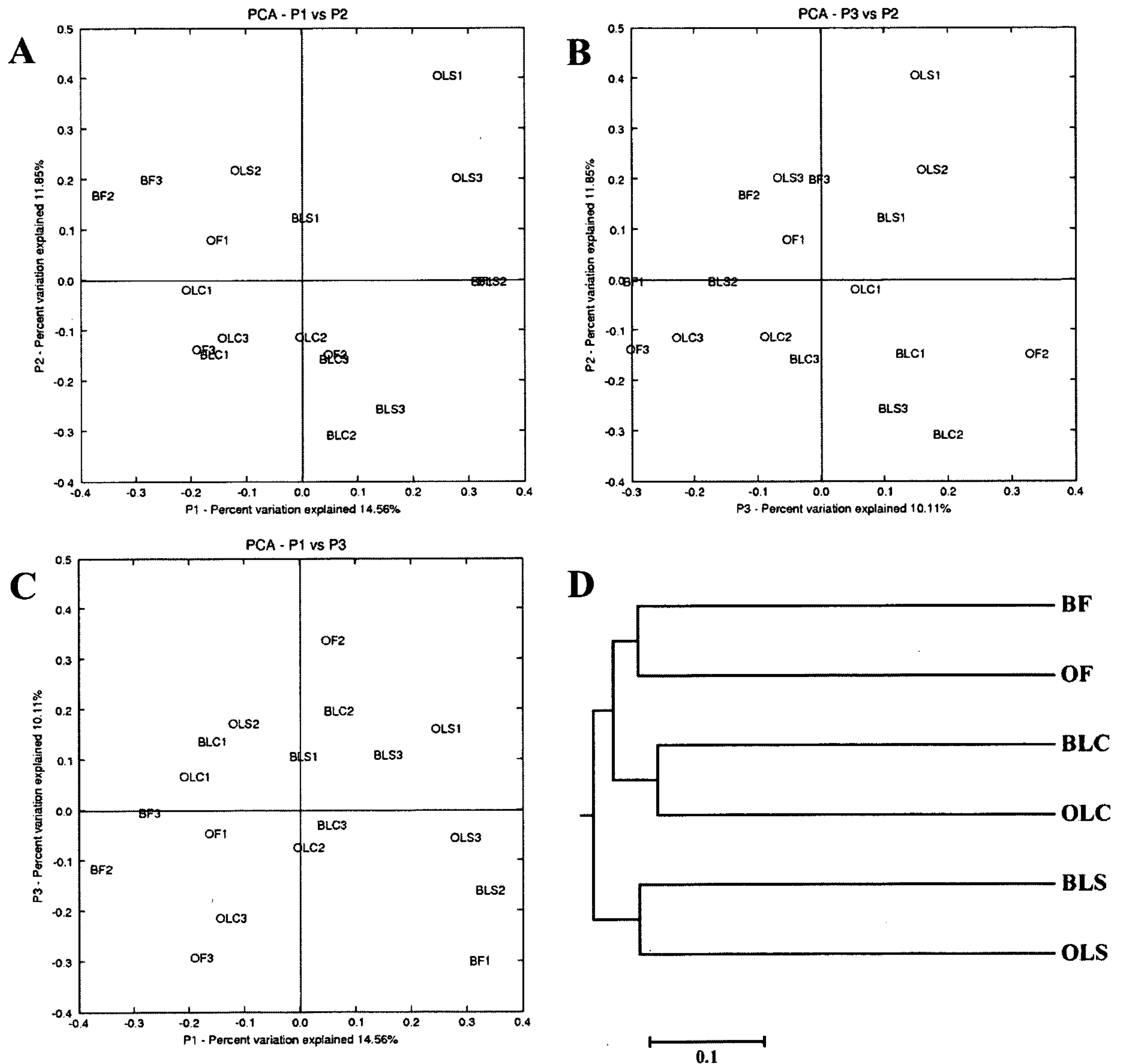


Figure 2.3: Whole community analysis for individual samples and pooled samples. Sample identification indicated by the first letter indicating color orange (O) or brown (B), followed by location Florida Keys (F), Little Cayman (LC), or Lee Stocking Island (LS), with replicate samples indicated by a number 1-3. Graphs depict the first three principal component analysis of individual replicate samples based on UniFrac distance calculations (A-C). Clustering of pooled communities by location and color based on UniFrac distance calculations. UMPGA based on UniFrac distance estimates for pooled replicates by location and color (D).

location and color they cluster together by location rather than color (D, Fig 2.3). However, differences in the pooled groups reflected in the clustering were not significant as determined by the UniFrac significance test (Data not shown).

Discussion

From clone libraries of the *nifH* gene we were able to compare the communities of bacteria potentially capable of fixing nitrogen associated with the Caribbean coral *Montastraea cavernosa*. Our goal was to identify differences between orange and brown color morphs that could be accredited for the previously documented difference in fixation between the two (Lesser et al. 2004; Lesser et al. 2007). Additionally we were interested in whether these differences were consistent for colonies in transition from brown to orange by sampling the transition area and comparing the nitrogen fixing bacterial communities between orange, transition, and brown polyps.

For the overall comparison of the orange versus brown communities we were interested in differences in specific OTUs for the whole community. For our OTU grouping we set the sequence similarity cutoff for 95%, a cut off commonly used in *nifH* diversity studies, which represents approximately species to genus taxonomic groupings (Konstantinidis 2005; Hsu and Buckley 2008). For specific OTUs several were found statistically different when comparing all orange versus brown colonies with most being more abundant in orange colonies. While no OTU representing cyanobacteria was found in association with all the orange colonies for orange versus brown colonies as found in previous studies (Lesser et al. 2004), the cyanobacterial OTU 17 had a greater abundance

in orange morphs in all clone libraries for all locations (Fig 2.1). Interestingly, a number of γ-proteobacterial OTUs were found in a greater abundance for orange versus brown colonies. Previous studies identified cyanobacteria as requiring a co-culture of heterotrophic bacteria in order to fix nitrogen, leading one to hypothesize that the presence of both the proteobacteria and the cyanobacteria may be required for fixation in *M. cavernosa* as well (Li et al. 2010). When comparing the abundance of specific OTU's between the individual polyps for the colonies in transition from brown to orange the trend of a greater abundance of OTU17 was not observed. However, the γ-proteobacteria OTU 35 and 37 were both found in a greater abundance in the orange versus transition and brown polyps. Though no clear dissimilarity in the communities as a whole or overall was observed that could account for the difference in fixation between the two morphs observed trends indicate specific OTUs may be responsible for this difference.

Alternatively, the difference in fixation could be due to differences in the community structure of the nitrogen fixing bacteria associated with the brown versus orange colonies. However, when comparing the community structure of the individual samples no observable trend was identified for either color or location (A-C, Fig 2.3). When treating the sequences from a particular location and color as a single sample they clustered by location rather than color (D, Fig 2.3). While the community structure for the replicates was more similar to each other based on location rather than color, the communities were all statistically similar. These results indicate that the structure of the communities as a whole could not account for the difference in fixation between the orange and brown color morphs.

There are three possible explanations for the lack of a universally present group of cyanobacteria associated with the orange colonies. Two are based on experimental biases. The first is PCR bias, and while this protocol was optimized as highly universal and was theorized to amplify cyanobacteria as well as proteobacteria, the potential for bias toward proteobacterial versus cyanobacteria *nifH* gene sequences has not been evaluated (See Chapter I). A second explanation is that recent work on coral associated bacteria revealed that immediate extraction of DNA versus freezing of samples upon collection results in a greater efficiency in extraction of cyanobacterial DNA (Sekar et al. 2009). Since our samples were frozen prior to extraction this could explain the low abundance of cyanobacteria *nifH* genes in our clone libraries. A third explanation is that the abundant and universally present endosymbiotic cyanobacteria are not responsible for difference in fixation between the two color morphs. In order to identify who is fixing the nitrogen *nifH* diversity must be analyzed at the expression level. Mohamed et al. (2008) investigated the diversity of nitrogen fixing bacteria associated with coral reef sponges. They found a similar diversity of nitrogen fixing bacteria associated with the sponge, but at the expression levels, transcripts were primarily from cyanobacteria. They noted that this lack of diversity at the expression level could be due to the time of sampling as *nifH* expression is known to exhibit diel patterns (Hewson et al. 2007).

One result of note is the great diversity of nitrogen fixing bacteria associated with both color morphs. While nitrogen is a known factor limiting productivity in coral reef ecosystems, why are only the bacteria associated with the orange color morph fixing nitrogen when bacteria capable of fixing nitrogen are associated with both morphs?

Could the difference in fixation be due to host control of the microbial community where the brown color morphs inhibit nitrogen fixation, and if so why?

The role of nitrogen fixing symbionts in the coral holobiont is largely unknown however; the application of diversity and physiology studies to this system provides significantly greater insight than either study individually. Furthermore as research to date indicates the coral nitrogen fixing bacteria symbiosis varies greatly between coral species and thus requires investigations of multiple coral species in order to fully understand the functional role of these microorganisms within the holobiont.

CHAPTER III

CONCLUSION

Anthropogenic factors have, and continue to, negatively impact the environment. Two such factors are global warming and eutrophication. The response of the nitrogen cycle to these factors determines the extent of their influence on both ecosystem function and services around the world (Vitousek et al. 1997b; Gruber and Galloway 2008). By examining how nitrogen cycling, and in particular nitrogen fixation, is predicted to respond to these disturbances a contextual framework can be created to address these two issues. For this conclusion the broad contextual implications of the data obtained in this thesis will be discussed as well as potential directions for future research.

Global warming is caused by increased atmospheric CO_2 levels and therefore directly affects the global carbon cycle (Cox et al. 2000). The global carbon cycle is also closely related to the global nitrogen cycle; one example is that atmospheric levels of NO_2 and rates of denitrification are significantly correlated with atmospheric CO_2 levels (Gruber and Galloway 2008). Because of the close coupling between the global carbon and nitrogen pathways and budgets it is fundamental to our ability to predict the affects of increased atmospheric CO_2 that scientist obtain an accurate global nitrogen budget (Falkowski 1997). The global nitrogen budget is estimated as either balanced or with a large deficit (Gruber and Galloway 2008). This discrepancy is based on significantly higher estimated rates of denitrification (Codispoti et al. 2007). Whether these budgets are balanced or not they are based on large uncertainties, one of which is the contribution

of nitrogen fixing bacteria to the global nitrogen cycle (Gruber and Galloway 2008). In order to decrease this level of uncertainty in the contribution of new nitrogen by nitrogen fixation researchers must reevaluate the diversity of nitrogen fixing bacteria in all ecosystems, including the world's oceans. Within the last 10 years a new group of ubiquitous and abundant pelagic unicellular nitrogen fixing cyanobacteria have been discovered (Zehr et al. 2001), and their measured rates of nitrogen fixation has allowed researchers to more accurately estimate the global contribution by nitrogen fixation to the marine nitrogen budget (Ward et al. 2007). Based on the evaluation of PCR primers for the *nifH* gene (see Chapter I), the most commonly used technique for identifying the diversity of bacteria in an environment; commonly used protocols may potentially miss some of the most abundant groups of nitrogen fixing bacteria, particularly those belonging to the classes δ- and α-proteobacteria (Chapter I). The use of the modified nested PCR protocol discussed in Chapter I could increase the identification of new groups of previously unidentified nitrogen fixing bacteria; potentially decreasing the level of uncertainty in the estimated contribution of nitrogen fixation to the marine nitrogen budget, and therefore to the global nitrogen budget.

Future research should continue to evaluate the role of PCR bias and primer mismatch in *nifH* diversity studies. Using an *in silico* approach the next step in analyzing published *nifH* primers is to determine the position of mismatches along the primer sequence. This additional analysis will help to predict the extent these mismatches affect any bias observed in these reactions. The role of bias should also be more thoroughly investigated in the laboratory setting. With new sequencing technologies e.g., 454

sequencing, researchers are able to obtain thousands of reads per sample. These new technologies have yet to be applied to *nifH* diversity studies and have the potential to revolutionize our understanding of these communities. With high read numbers comes greater sample coverage. With this new technology it is now possible to analyze the *nifH* diversity for a single sample with all primer sets analyzed in this study. Based on these results one could determine the role the predicted mismatches played in determining the diversity of *nifH* genes recovered. While this new technology could serve to ensure more complete coverage the extent of PCR bias still remains unknown. To identify bias and potentially optimize protocols to reduce bias two approaches should be utilized. First is to analyze samples of known diversity. Using cultures of nitrogen fixing bacteria, researches can produce a mixed sample with known proportions. They could then relate back the ratios of sequence products to determine the extent of PCR bias. A second approach is to utilize quantitative PCR to determine the abundance of specific template sequences in a sample. You could then compare the abundances for multiple template sequences to one another to the ratios obtained while analyzing the sequences using high throughput technologies. The results of these studies would provide the ground work for the development of *nifH* diversity assays that would potentially reveal a greater diversity of nitrogen fixing bacteria as well as provide a more accurate representation of the community structure.

A number of highly productive ecosystems such as coral reefs and tropical rain forests have successfully adapted to and thrive in nutrient poor environments, and the addition of excess nutrients via eutrophication has the potential to disrupt the balance of

these and other delicate ecosystems (Vitousek et al. 1997b). Elevated nitrogen levels caused by terrestrial runoff from agriculture and poorly treated sewage cause productivity in traditionally oligotrophic systems to no longer be limited by biologically available nitrogen (Vitousek et al. 1997b). With nitrogen fixation no longer required to supply new nitrogen for primary productivity and balanced growth in evolutionarily nitrogen poor environments, what affect does this have on the community of nitrogen fixing bacteria present within an ecosystem and how does this change their role within the system? While excess nitrogen can increase the productivity of an ecosystem, it also disrupts the balance of the system, by eliminating bottom up control (Dinsdale et al. 2008). In other words organisms whose growth is normally limited by low nitrogen levels is unchecked and able to out compete other organisms for resources such are space and sunlight.

Coral reefs evolved as highly productive ecosystems in nutrient poor waters (Webb et al. 1975). This high productivity was accredited to the close coupling and efficient cycling of organic nitrogen mineralization and inorganic nitrogen utilization as well as the presence of nitrogen fixing bacteria providing "new" nitrogen into the system (Wiebe et al. 1975). However, the productivity of reefs neighboring highly populated coastal areas are no longer nitrogen limited due to terrestrial nutrient runoff. This excess nutrient disrupts the balance of the systems (Dinsdale et al. 2008, Sandin et al. 2008). The input of excess biologically available nitrogen into a system favors organisms that are dependent upon high nitrogen levels versus those that have adapted to life styles favoring nitrogen poor environments, in turn reducing the biodiversity within the system (Vitousek et al. 1997b). In particular the addition of biologically available nitrogen

undermines the role of nitrogen fixing bacteria in these systems. Recent research has identified a diverse and ubiquitous community of nitrogen fixing bacteria associated with reef building corals (Chapter II,Olson et al. 2009; Kimes et al. 2010). However, nitrogen fixation was not observed for all corals with nitrogen fixing bacteria associated with them, namely the brown color morph of the Caribbean coral *Montastraea cavernosa* (Chapter II). Based on whole community analysis of bacteria potentially capable of fixing nitrogen associated with *Montastraea cavernosa* there was a trend reflecting similarities among location versus color although these differences were not significant (Chapter II). Lee Stocking Island and Little Cayman Island have similarly lows levels of anthropogenic impact as they are nowhere near large population center and have sub micromolar water nutrient concentrations. Conversely, Florida Keys are well documented for their high impact levels primarily due their close proximity to large population centers. Interestingly, the community of nitrogen fixing bacteria was more similar between the Little Cayman colonies and Florida Keys, than Lee Stocking Island. These results suggest that location and not the level of anthropogenic impact plays a larger role in the overall community structure. So while the community composition was more dependent upon location than level of impact, on a regional scale does the community of nitrogen fixing bacteria change with increased levels of nutrients present within a system? Also, how does the role of the nitrogen fixing bacteria change with increased levels of biologically available nitrogen? Bacterial species of the genus *Vibrio* have been identified as the pathogenic agent for a number of coral diseases (Ben-Haim et al. 1999; Ben et al. 2002), while at the same time other *Vibrio* species were identified as

potential nitrogen fixing symbionts (Chimetto et al. 2008; Olson et al. 2009). Though there is no evidence that the symbiotic nitrogen fixing bacteria are pathogenic or that the pathogenic species are capable of nitrogen fixation, could beneficial symbiotic nitrogen fixing bacteria become opportunistic pathogens in the coral system once nitrogen fixation is no longer required? With increased nitrogen levels their growth could be unchecked leading them to be a causative agent of disease. It is important to compare the role of these coral symbionts in nitrogen poor and nutrient rich systems to see how they change. With a greater understanding of the role of these symbionts in coral productivity and potentially disease we can contribute to conservation efforts. This can be done in two ways. 1) For the restoration of coral reefs where new corals are transplanted into oligotrophic areas, coral species known to host nitrogen fixing bacteria that actively fix nitrogen such as orange color morphs of *Montastraea cavernosa* maybe a key player in the successful restoration of a reef. This would function in a similar fashion to crop rotations where legumes, plants with symbiotic nitrogen fixing bacteria, are used to fertilize agricultural systems. Additionally, for corals that may be dependent on nitrogen fixing symbionts in nitrogen poor environments it is important to monitor their health and stress levels in nutrient rich systems. The results from these analyses may provide additional sound scientific findings to support efforts aimed at decreasing nutrient runoff to these heavily impacted reef ecosystems.

At this point the role of nitrogen fixing bacteria in the coral holobiont is largely unknown. For *Montastraea cavernosa* the most well understood system, major questions remain unanswered. One of which is; what is the cause for the difference in fixation

between the orange and brown color morphs. Results presented in this thesis identified microorganisms potentially responsible for this difference namely groups of cyanobacteria and proteobacteria (Chapter II). The first step in identifying these as the microorganisms responsible for the difference in fixation between the two color morphs is to look at *nifH* diversity at the expression level. The results of this study will determine which organisms are actively fixing nitrogen. Furthermore the expression should be analyzed over a diel period as previous studies have indicated that fixation varies over a 24 hr period with peaks in fixation at dawn and dusk. Once candidate microorganisms are identified quantitative PCR assays should be applied in order to quantify potential differences in symbiont abundance between the two color morphs. Additionally, the location of these potential microorganisms in relation to the coral host should be investigated. Two approaches can be taken to address this. The first of which is to use fluorescent *in situ* hybridization. Using probes designed specifically for the *nifH* gene combined with laser confocal microscopy researchers will be able to determine the microorganism's location in relation to the host. A second approach would utilize a new technology multisotope mass spectroscopy (MIMS). This analysis looks at the isotopic composition of a sample at nanometer resolution. When combing a N^{15} pulse chase isotope experimental design and *in* situ hybridization for *nifH* genes of interest the results of this analysis would reveal not only the identity of the nitrogen fixing symbiont but also whether the product of fixation is utilized by the coral host and/or algal symbiont. The results of these potential studies will help to elucidate the importance of the coral

nitrogen fixing bacterial symbiosis for the coral holobiont and coral reef ecosystems as a whole.

Global warming and nutrient runoff are well documented as having a serious detrimental effect on global ecosystems and coral reefs in particular (Carpenter et al. 2008; Dinsdale et al. 2008). While identifying the diversity of bacteria potentially capable of fixing nitrogen is not going to resolve either of these pressing issues it can make an important contribution into understanding the full effect of these anthropogenic factors.

REFERENCES

Arrigo KR (2005) Marine microorganisms and global nutrient cycles. Nature 437:349-355

Belnap J (2002) Nitrogen fixation in biological soil crusts from southeast Utah, USA. Biology and Fertility of Soils 35:128-135

Ben-Haim Y, Banim E, Kushmaro A, Loya Y, Rosenberg E (1999) Inhibition of photosynthesis and bleaching of zooxanthellae by the coral pathogen *Vibrio shiloi*. Envrionmental Microbiology 1:223-229

Ben H, Ben-Haim Y, Rosenberg, Rosenberg E (2002) A novel *Vibrio* sp. pathogen of the coral *Pocillopora damicornis*. Marine Biology 141:47-55

Burgess B (2002) The mechanism of molybdenum nitrogenase: an overview. Nitrogen Fixation: From Molecules to Crop Productivity, pp13-18

Cantera JJL, Kawasaki H, Seki T (2004) The nitrogen-fixing gene (*nifH*) of *Rhodopseudomonas palustris*: a case of lateral gene transfer? MICROBIOLOGY 150:2237-2246

Capone DG, Carpenter EJ (1982) Nitrogen-fixation in the marine-environment. Science 217:1140-1142

Capone DG, Zehr, J.P., Paerl, H.W., Bergman, B. Carpenter, E.J. (1997) *Trichodesmium*, a globally significant marine cyanobacterium. Science 276:1226-1229

Carpenter KE, Abrar M, Aeby G, Aronson RB, Banks S, Bruckner A, Chiriboga A, Cortes J, Delbeek JC, DeVantier L, Edgar GJ, Edwards AJ, Fenner D, Guzman HM, Hoeksema BW, Hodgson G, Johan O, Licuanan WY, Livingstone SR, Lovell ER, Moore JA, Obura DO, Ochavillo D, Polidoro BA, Precht WF, Quibilan MC, Reboton C, Richards ZT, Rogers AD, Sanciangco J, Sheppard A, Sheppard C, Smith J, Stuart S, Turak E, Veron JEN, Wallace C, Weil E, Wood E (2008) One-third of reef-building corals face elevated extinction risk from climate change and local impacts. Science 321:560-563

Charpy L, Alliod R, Rodier M, Golubic S (2007) Benthic nitrogen fixation in the SW New Caledonia lagoon. Aquatic Microbial Ecology 47:73-81

Chimetto LA, Brocchi M, Thompson CC, Martins RCR, Ramos HR, Thompson FL (2008) Vibrios dominate as culturable nitrogen-fixing bacteria of the Brazilian coral *Mussismilia hispida*. Systematic and Applied Microbiology 31:312-319

Church MJ, Mahaffey C, Letelier RM, Lukas R, Zehr JP, Karl DM (2009) Physical forcing of nitrogen fixation and diazotroph community structure in the North Pacific subtropical gyre. Global Biogeochemical Cycles 23:-

Cox PM, Betts RA, Jones CD, Spall SA, Totterdell IJ (2000) Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. Nature 408:184-187

Crews TE, Kurina LM, Vitousek PM (2001) Organic matter and nitrogen accumulation and nitrogen fixation during early ecosystem development in Hawaii. Biogeochemistry 52:259-279

Crossland CJ, Barnes DJ (1976) Acetylene reduction by coral skeletons. Limnology and Oceanography 21:153-156

Diallo MD, Reinhold-Hurek B, Hurek T (2008) Evaluation of PCR primers for universal *nifH* gene targeting and for assessment of transcribed *nifH* pools in roots of *Oryza longistaminata* with and without low nitrogen input. Fems Microbiology Ecology 65:220-228

Dinsdale EA, Pantos O, Smriga S, Edwards RA, Angly F, Wegley L, Hatay M, Hall D, Brown E, Haynes M, Krause L, Sala E, Sandin SA, Thurber RV, Willis BL, Azam F, Knowlton N, Rohwer F (2008) Microbial ecology of four coral atolls in the Northern Line Islands. PLoS ONE 3:e1584

Distel DL, Beaudoin DJ, Morrill W (2002) Coexistence of multiple proteobacterial endosymbionts in the gills of the wood-boring bivalve *Lyrodus pedicellatus* (Bivalvia : Teredinidae). Applied and Environmental Microbiology 68:6292-6299

Dong ZB, Wang XM, Liu LY (2000) Wind erosion in arid and semiarid China: an overview. Journal of Soil and Water Conservation 55:439-444

Douglas AE (1995) The ecology of symbiotic micro-organisms. In: Begon M, Fitter AH (eds) Advances in Ecological Research. Academic Press, London, pp69-95

Eskew DL, Ting IP (1978) Nitrogen fixation by legumes and blue-green algal-lichen crusts in a Colorado desert environment. American Journal of Botany:850-856

Falkowski PG (1997) Evolution of the nitrogen cycle and its influence on the biological sequestration of CO_2 in the ocean. Nature 387:272-275

Fanning KA (1989) Influence of atmospheric-pollution on nutrient limitation in the ocean. Nature 339:460-463

Fedorov DN, Ivanova EG, Doronina NV, Trotsenko YA (2008) A new system of degenerate oligonucleotide primers for detection and amplification of nifHD genes. MICROBIOLOGY 77:247-249

France SC, Kocher TD (1996) DNA sequencing of formalin fixed crustaceans from archival research collections. Molecular Marine Biology and Biotechnology 5:304-313

Galloway JN, Schlesinger WH, Levy H, Michaels A, Schnoor JL (1995) Nitrogen-fixation - anthropogenic enhancement-environmental response. Global Biogeochemical Cycles 9:235-252

Gruber N, Galloway JN (2008) An Earth-system perspective of the global nitrogen cycle. Nature 451:293-296

Hamner WM (1995) Predation, cover, and convergent evolution in epipelagic oceans. Marine and Freshwater Behavior and Physiology 26:71-89

Hamner WM, Colin PL, Hamner PP (2007) Export-import dynamics of zooplankton on a coral reef in Palau. Marine Ecological Progress Series 334:83-92

Hewson I, Moisander PH, Morrison AE, Zehr JP (2007) Diazotrophic bacterioplankton in a coral reef lagoon: phylogeny, diel nitrogenase expression and response to phosphate enrichment. The ISME Journal 1:78-91

Hsu S-F, Buckley DH (2008) Evidence for the functional significance of diazotroph community structure in soil. Isme Journal 3:124-136

Ishii K, Fukui M (2001) Optimization of annealing temperature to reduce bias caused by a primer mismatch in multitemplate PCR. Applied Enviromental Microbiology 67:3753-3755

Kanagawa T (2003) Bias and artifacts in multitemplate polymerase chain reactions (PCR). Journal of Bioscience and Bioengineering 96:317-323

Karl D, Michaels A, Bergman B, Capone D, Carpenter E, Letelier R, Lipschultz F, Paerl H, Sigman D, Stal L (2002) Dinitrogen fixation in the world's oceans. Biogeochemistry (Dordrecht) 57-58:47-98

Kimes NE, Van Nostrand JD, Weil E, Zhou J, Morris PJ (2010) Microbial functional structure ofMontastraea faveolata, an important Caribbean reef-building coral, differs between healthy and yellow-band diseased colonies. Environmental Microbiology 12:541-556

Klausmeier CA, Litchman E, Levin SA (2004) Phytoplankton growth and stoichiometry under multiple nutrient limitation. Limnology and Oceanography 49:1463-1470

Kneip C, Lockhart P, Voß C, Maier U-G (2007) Nitrogen fixation in eukaryotes – New models for symbiosis. BMC Evolutionary Biology 7:55

Knowlton N, Rohwer F (2003) Multispecies microbial mutualisms on coral reefs: The host as a habitat. Am Nat 162:S51-S62

Konstantinidis KT (2005) Genomic insights that advance the species definition for prokaryotes. Proceedings of the National Academy of Sciences 102:2567-2572

Kurina LM, Vitousek PM (2001) Nitrogen fixation rates of *Stereocaulon vulcani* on young Hawaiian lava flows. Biogeochemistry 55:179-194

Larkum AWD, Kennedy IR, Muller WJ (1988) Nitrogen fixation on a coral reef. Marine Biology 98:143-155

Lesser MP, Mazel CH, Gorbunov MY, Falkowski PG (2004) Discovery of Symbiotic Nitrogen-Fixing Cyanobacteria in Corals. Science 305:997-1000

Lesser MP, Falcón LI, Rodríguez-Román A, Enríquez S, Hoegh-Guldberg O, Iglesias-Prieto R (2007) Nitrogen fixation by symbiotic cyanobacteria provides a source of nitrogen for the scleractinian coral Montastraea cavernosa. Marine Ecology Progress Series 346:143-152

Li QP, Hansell DA (2008) Nutrient distributions in baroclinic eddies of the oligotrophic North Atlantic and inferred impacts on biology. Deep-Sea Research Part Ii-Topical Studies in Oceanography 55:1291-1299

Li Z, Yu J, Kim K-R, Brand J (2010) Nitrogen fixation by a marine non-heterocystous cyanobacterium requires a heterotrophic bacterial consort. Environmental Microbiology 12:1185-1193

Mårtensson L, Díez B, Wartiainen I, Zheng W, El-Shehawy R, Rasmussen U (2009) Diazotrophic diversity, *nifH* gene expression and nitrogenase activity in a rice paddy field in Fujian, China. Plant and Soil 325:207-218

Marusina AI, Boulygina ES, Kuznetsov BB, Tourova TP, Kravchenko IK, Gal'chenko VF (2001) A system of oligonucleotide primers for the amplification of *nifH* genes of different taxonomic groups of prokaryotes. MICROBIOLOGY 70:73-78

Mehta MP, Butterfield DA, Baross JA (2003) Phylogenetic Diversity of Nitrogenase (nifH) Genes in Deep-Sea and Hydrothermal Vent Environments of the Juan de Fuca Ridge. Applied and Environmental Microbiology 69:960-970

Mohamed NM, Colman AS, Tal Y, Hill RT (2008) Diversity and expression of nitrogen fixation genes in bacterial symbionts of marine sponges. Environmental Microbiology 10:2910

O'Neil JM, Capone DG (2008) Nitrogen cycling in coral reef envrionments. In: Capone DG, Bronk DA, Mulholland MR, Carpenter EJ (eds) Nitrogen in the Marine Environment. Elsevier, pp949-989

Odintsov VS (1991) Nitrogen fixation in *Prochloron* (Prochlorophyta)- ascidian assoications: is procholorn responsible? Endocytobiosis and Cell Research 7:253-258

Odintsov VS, Propp MV, Voskoboynikov GM (1987) Nitrogen fixation and symbionts assoicated with hydrocoral millepora. Endocytobiosis and Cell Research 4:197-203

Ohkuma M, Kudo T (1996) Phylogenetic diversity of the intestinal bacterial community in the termite *Reticulitermes speratus*. Appl Environ Microbiol 62:461-468

Olson ND, Ainsworth TD, Gates RD, Takabayashi M (2009) Diazotrophic bacteria associated with Hawaiian *Montipora* corals: Diversity and abundance in correlation with symbiotic dinoflagellates. Journal of Experimental Marine Biology and Ecology 371:140-146

Paerl HW (1984) Nitrogen fixation nitrogenase activity attributable to a specific prochloron prochlorophyta-ascidian association in Palau Micronesia. Marine Biology (Berlin) 81:251-254

Poly F, Monrozier LJ, Bally R (2001) Improvement in the RFLP procedure for studying the diversity of *nifH* genes in communities of nitrogen fixers in soil. Research in Microbiology 152:95-103

Polz MF, Cavanaugh CM (1998) Bias in template-to-product ratios in multitemplate PCR. Applied Enviromental Microbiology 64:3724-3730

Polz MF, Disrtel, D.L., Zarda, B., Amann, R., Felbeck, H. Ott, J.A. Cavanaugh, C. (1994) Phylogenetic analysis of a highly specific association between ectosymbiotic, sulfur-oxidizing bacteria and a marine nematode. Applied Enviromental Microbiology 60:4461-4467

Prechtl J, Kneip C, Lockhart P, Wenderoth K, Maier UG (2004) Intracellular spheroid bodies of Rhopalodia gibba have nitrogen-fixing apparatus of cyanobacterial origin. Molecular Biology and Evolution 21:1477-1481

Quigg A, Finkel ZV, Irwin AJ, Rosenthal Y, Ho TY, Reinfelder JR, Schofield O, Morel FMM, Falkowski PG (2003) The evolutionary inheritance of elemental stoichiometry in marine phytoplankton. Nature 425:291-294

Raymond J, Siefert JL, Staples CR, Blankenship RE (2004) The natural history of nitrogen fixation. Molecular Biology and Evolution 21:541-554

Redfield AC (1958) The biological control of chemical factors in the environment. American Scientist Autumn 1958:205-221

Rohwer F, Seguritan V, Azam F, Knowlton N (2002) Diversity and distribution of coral-associated bacteria. Marine Ecology Progress Series 243:1-10

Sarita S, Priefer UB, Prell J, Sharma PK (2008) Diversity of *nifH* gene amplified from rhizosphere soil DNA. Current Science 94:7

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Applied Enviromental Microbiology 75:7537-7541

Sekar R, Kaczmarsky LT, Richardson LL (2009) Effect of freezing on PCR amplification of 16S rRNA genes from microbes associated with black band disease of corals. Applied Enviromental Microbiology 75:2581-2584

Seneviratne G, Indrasena IK (2006) Nitrogen fixation in lichens is important for improved rock weathering. Journal of Biosciences (Bangalore) 31:639-643

Seutin G, White BN, Boag PT (1991) Preservation of avian blood and tissue samples for DNA analysis. Canadian journal of zoology 69:82-90

Shashar N, Cohen Y, Loya Y, Sar N (1994) Nitrogen fixation (acetylene reduction) in stony corals: Evidence for coral-bacteria interactions. Marine Ecology Progress Series 111:259-264

Sipos R, Székely A, J. , Palatinszky M, Révész S, Márialigeti K, Nikolausz M (2007) Effect of primer mismatch, annealing temperature and PCR cycle number on 16S rRNA gene-targetting bacterial community analysis. Fems Microbiology Ecology 60:341-350

Staples CR, Lahiri S, Raymond J, Von Herbulis L, Mukhophadhyay B, Blankenship RE (2007) Expression and Association of Group IV Nitrogenase NifD and NifH Homologs in the Non-Nitrogen-Fixing Archaeon Methanocaldococcus jannaschii. Journal Of Bacteriology 189:7392-7398

Steward GF, Zehr JP, Jellison R, Montoya JP, Hollibaugh JT (2004) Vertical distribution of nitrogen-fixing phylotypes in a meromictic, hypersaline Lake. Microbial Ecology 47:30-40

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. Molecular Biology and Evolution 24:1596-1599

Tan Z, Hurek T, Reinhold-Hurek B (2003) Effect of N-fertilization, plant genotype and environmental conditions on *nifH* gene pools in roots of rice. Environmental Microbiology 5:1009-1015

Tripp HJ, Bench SR, Turk KA, Foster RA, Desany BA, Niazi F, Affourtit JP, Zehr JP (2010) Metabolic streamlining in an open-ocean nitrogen-fixing cyanobacterium. Nature 464:90-94

Tripplett EW (2000) Prokaryotic nitrogen fixation: A model system for the analysis of a biological process. Horizon Scientific Press, Norfolk

Ueda T, Suga Y, Yahiro N, Matsuguchi T (1995) Remarkable N_2-fixing bacterial diversity detected in rice roots by molecular evolutionary analysis of *nifH* gene sequences. Journal Of Bacteriology 177:1414-1417

Vitousek PM, Hattenschwiler S, Olander L, Allison S (1997a) Nitrogen and nature. Ambio 31:97-101

Vitousek PM, Aber JD, Howarth RW, Likens GE, Matson PA, Schindler DW, Schlesinger WH, Tilman DG (1997b) Human alteration of the global nitrogen cycle:sources and consequences. Ecological Applications 7:737-750

Wang S-Z, Chen J-S, Johnson JL (1988) The presence of five *nifH*-like seuqences in *Clostridium pasturianum*: sequences divergence and transcription properties. Nucleic Acids Research 16:439-454

Ward BB, Capone DG, Zehr JP (2007) What's new in the nitrogen cycle? Oceanography 20:101-109

Webb KL, DuPaul WD, Wiebe W, Sottile W, Johannes RE (1975) Enewetak (Eniwetok) Atoll: Aspects of the nitrogen cycle on a coral reef. Limnol Oceanogr 20:198-210

Wegley L, Edwards R, Rodriguez-Brito B, Liu H, Rohwer F (2007) Metagenomic analysis of the microbial community associated with the coral Porites astreoides. Environmental Microbiology 9:2707-2719

Widmer F, Shaffer BT, Porteous LA, Seidler RJ (1999) Analysis of *nifH* gene pool complexity in soil and litter at a douglas fir forest site in the Oregon Cascade Mountain range. Applied Enviromental Microbiology 65:374-380

Wiebe WJ, Johannes RE, Webb KL (1975) Nitrogen fixation in a coral reef community. Science 188:257-259

Williams WM, Viner AB, Broughton WJ (1987) Nitrogen fixation (acetylene reduction) associated with the living coral *Acropora variabilis*. Marine Biology 94:531-535

Young JPW (1992) Phylogenetic classification of nitrogen-fixing organisms. In: Stacey G, Burris RH (eds) Biological nitrogen fixation. Chapman and Hall, Inc., New York, NY, pp43-86

Zani S, Mellon MT, Collier JL, Zehr JP (2000) Expression of nifH Genes in Natural Microbial Assemblages in Lake George, New York, Detected by Reverse Transcriptase PCR. Appl Environ Microbiol 66:3119-3124

Zehr JP, McReynolds LA (1989) Use of degenerate oligonucleotides for amplification of the *nifH* gene from the marine cyanobacterium Trichodesmium thiebautii. Appl Environ Microbiol 55:2522-2526

Zehr JP, Capone DG (1996) Problems and promises of assaying the genetic potential for nitrogen fixation in the marine environment. Microbial Ecology 32:263-281

Zehr JP, Turner PJ (2001) Nitrogen fixation:nitrogenase gene and gene expression. In: Paul JH (ed) Marine Microbiology. Academic Press, San Diego,

Zehr JP, Jenkins BD, Short SM, Steward GF (2003) Nitrogenase gene diversity and microbial community structure: a cross-system comparison. Environmental Microbiology 5:539-554

Zehr JP, Waterbury JB, Turner PJ, Montoya JP, Omoregie E, Steward GF, Hansen A, Karl DM (2001) Unicellular cyanobacteria fix N_2 in the subtropical North Pacific Ocean. Nature 412:635-638

Zehr JP, Montoya JP, Jenkins BD, Hewson I, Mondragon E, Short CM, Church MJ, Hansen A, Karl DM (2007) Experiments linking nitrogenase gene expression to nitrogen fixation in the North Pacific subtropical gyre. Limnology and Oceanography 52:169-183

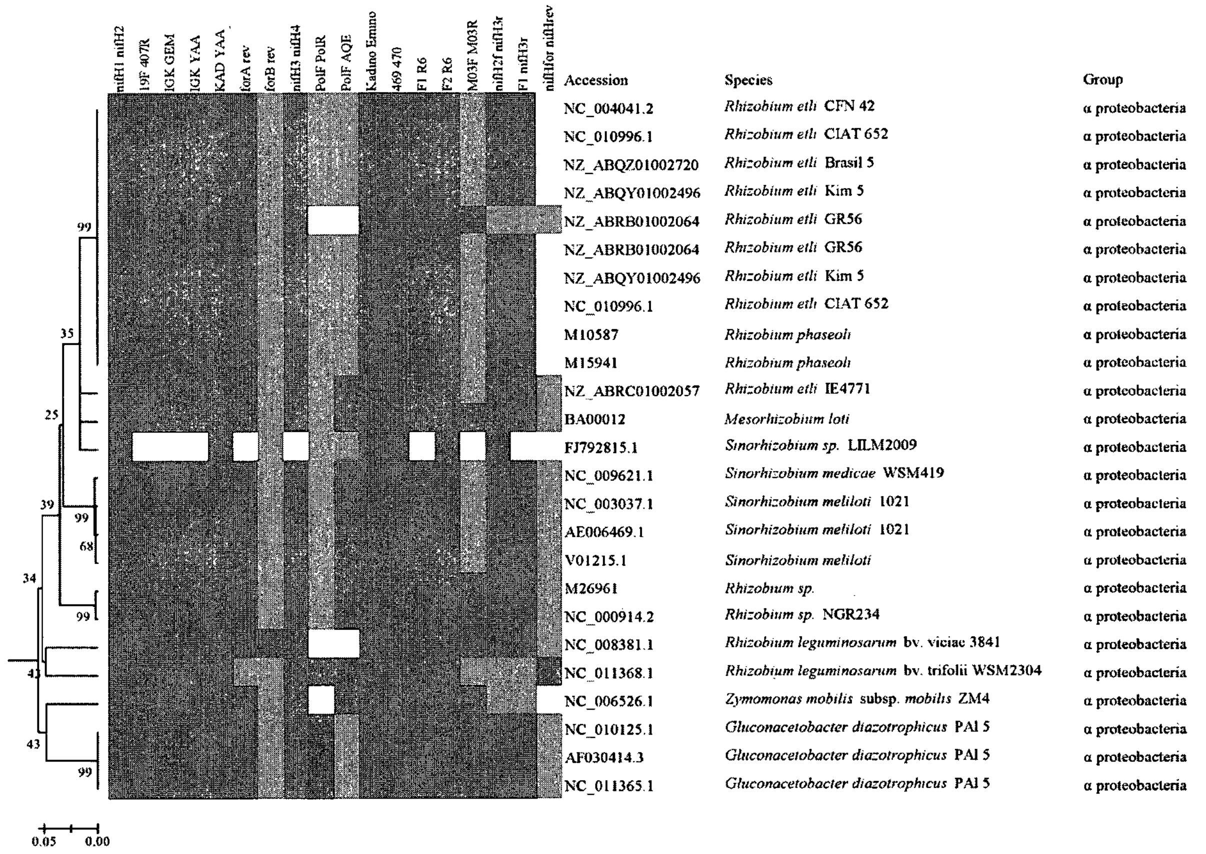
Zehr JP, Jenkins, B.D., Short, S.M., Steward, G.F. (2003) Nitrogenase gene diversity and microbial community structure: a cross-system comparison. Environmental Microbiology 5:539-554

Zehr JPWBB (2002) Nitrogen cycling in the ocean: New perspective on processes and paradigms. Applied and Environmental Microbiology 68:1015-1024

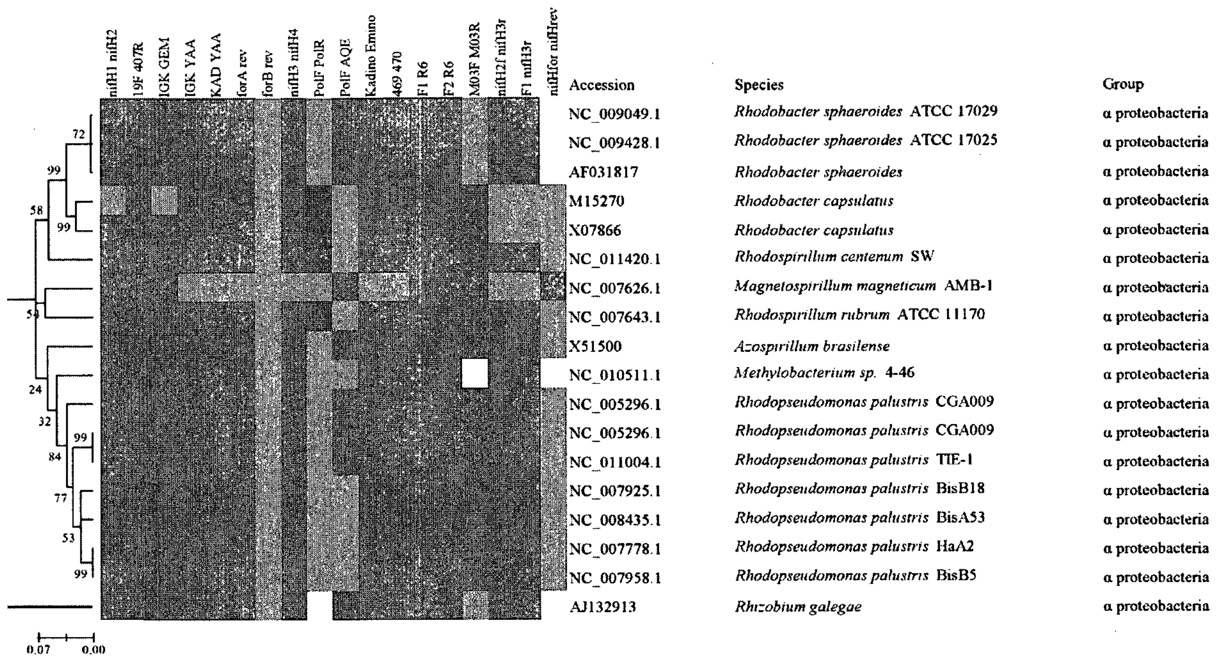
APPENDIX

Supplemental figures for Chapter I. Breakdown of primer pair mismatches to each sequence. Primer pairs are listed along the top of the figure and sequences accession number species name and group are on the left. The sequences are organized based on phylogeny with the groups sub tree indicated on the right with the scale bar indicating the substitution rate. How well each of the primers match-up with each of the sequences is indicated by the color of the intersecting blue (0 mismatches), green (1 or 2 mismatches), or red (3 or 4 mismatches). Each subgroup is represented in individual figures group 1.1 (A), 1.2(B), 1.3(C), 1.4(D), 1.5(E), 1.6(F), 2.1(G), 2.2(H), 3(I), 4(J).

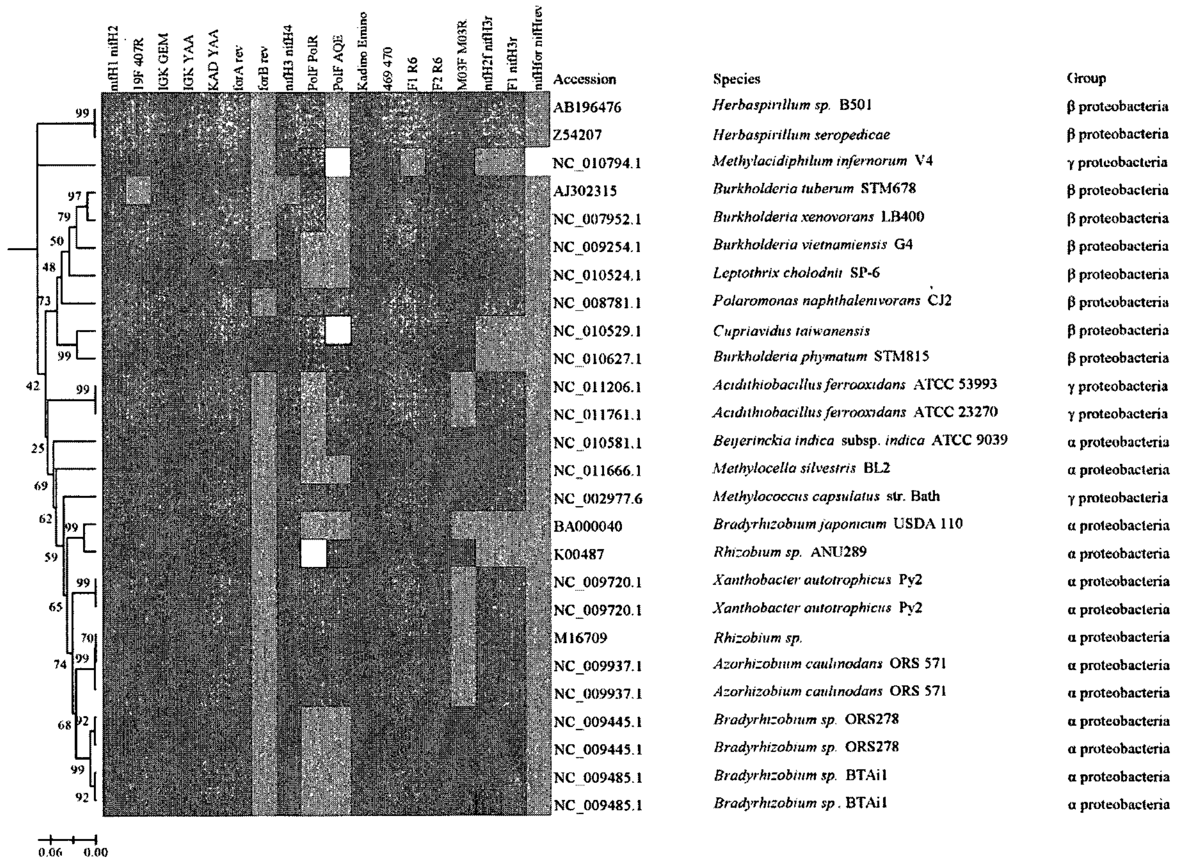
A



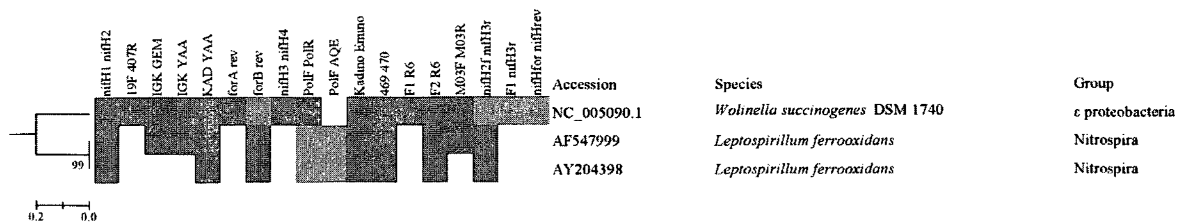
B



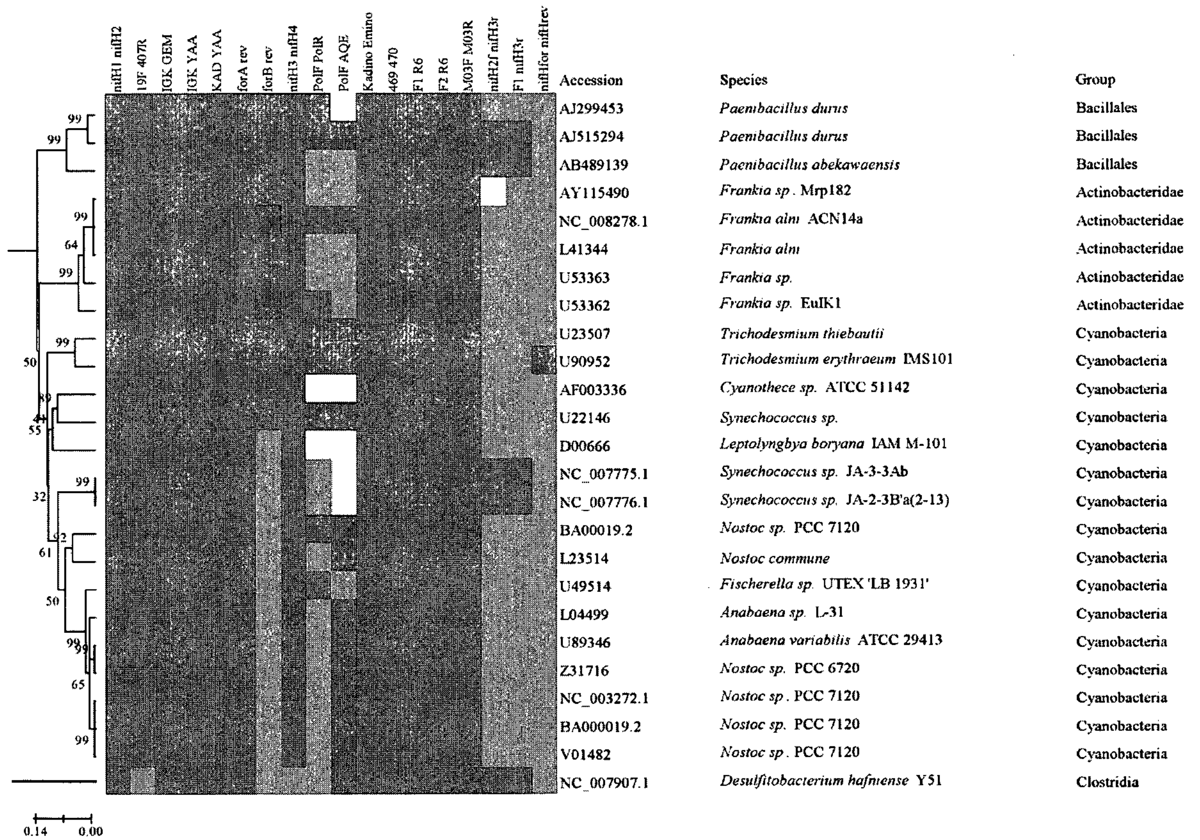
C



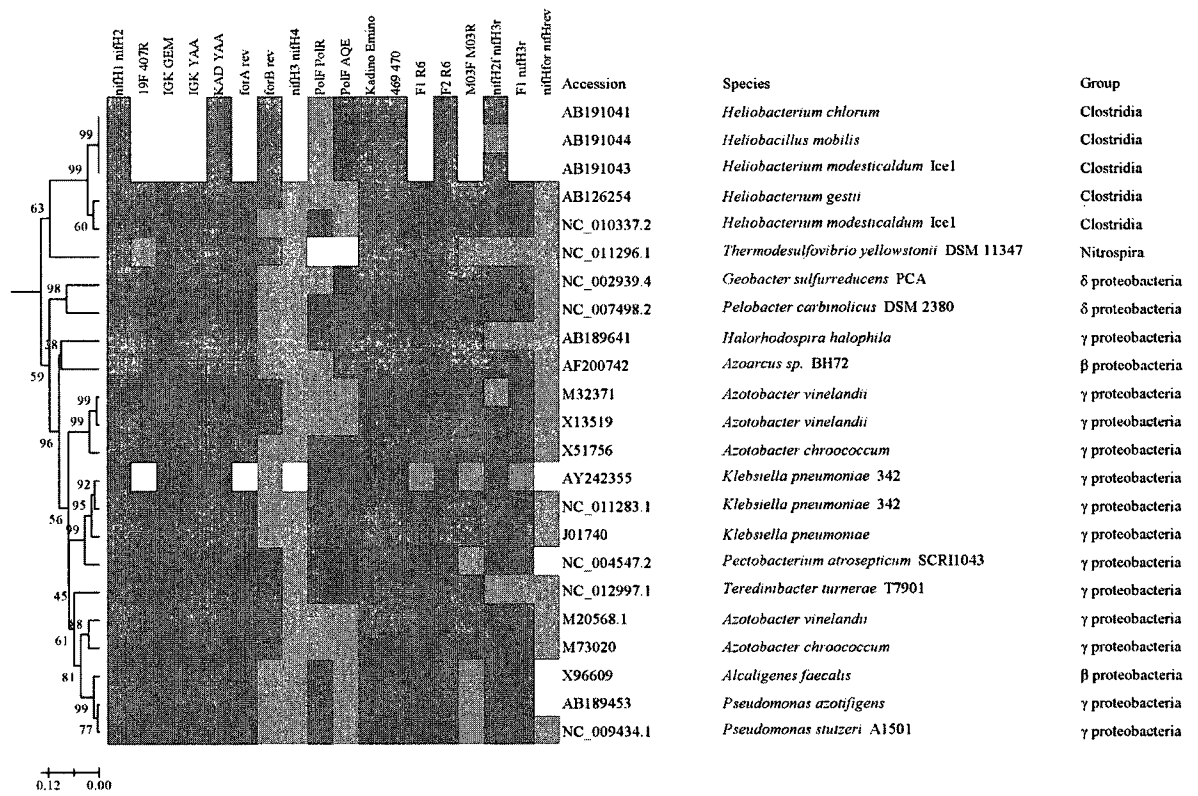
D



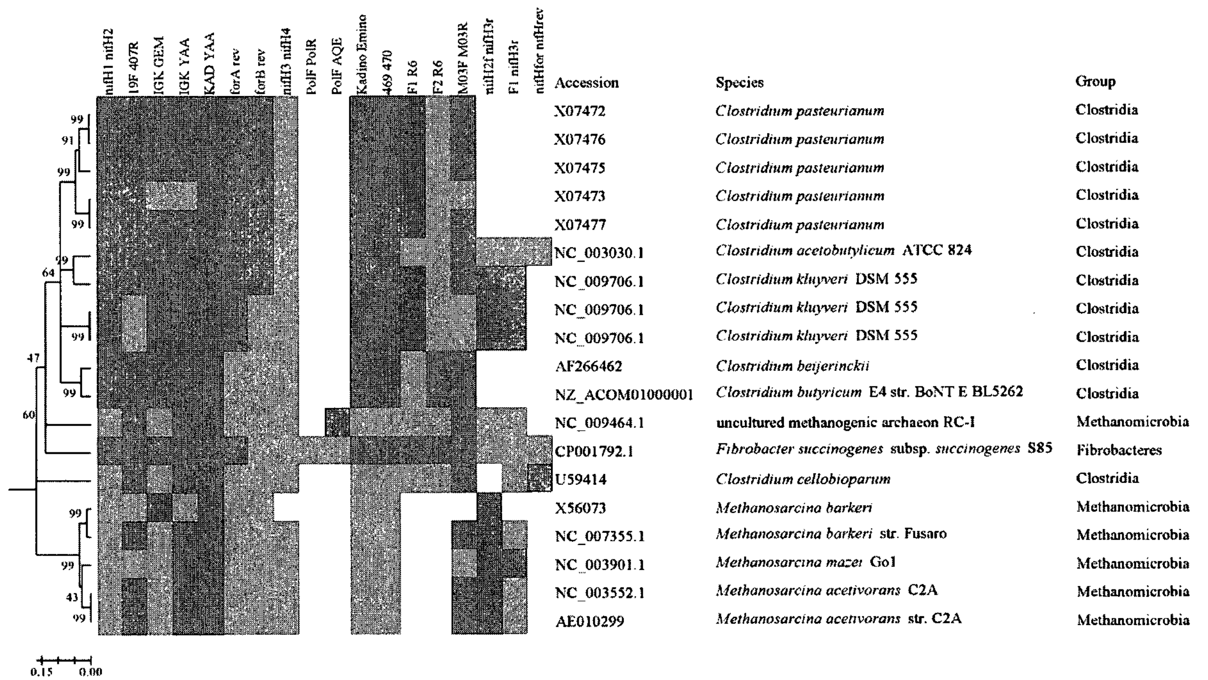
E



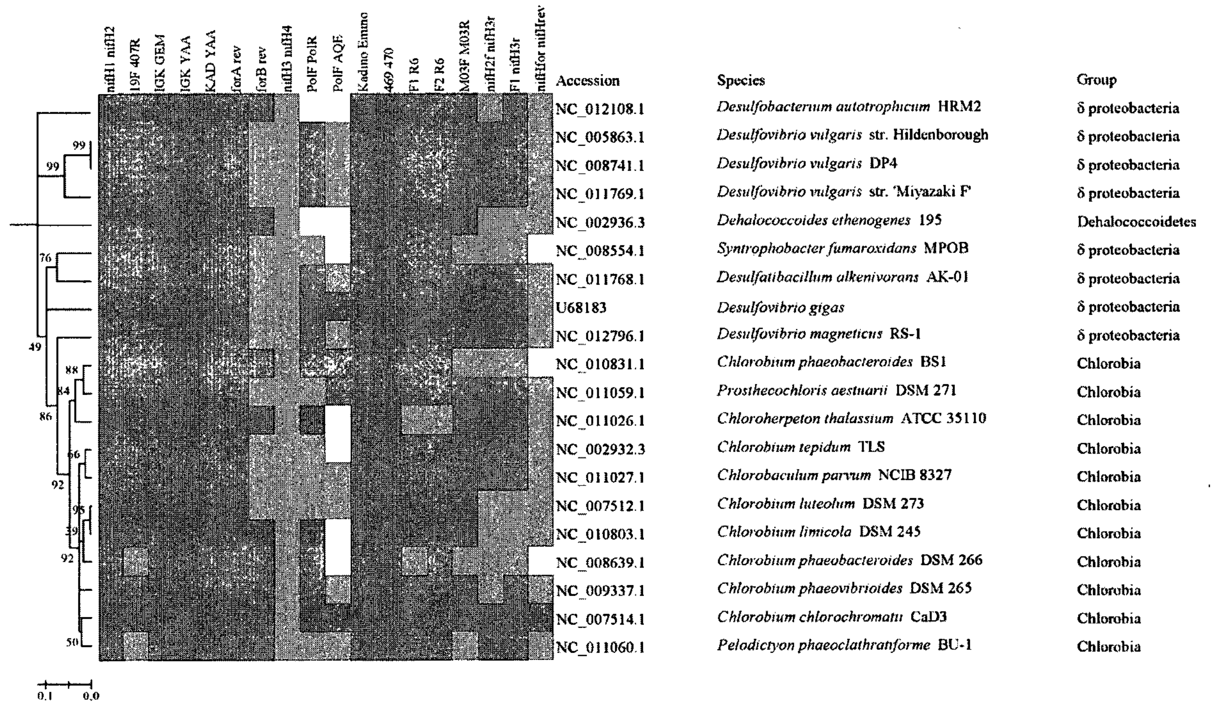
F



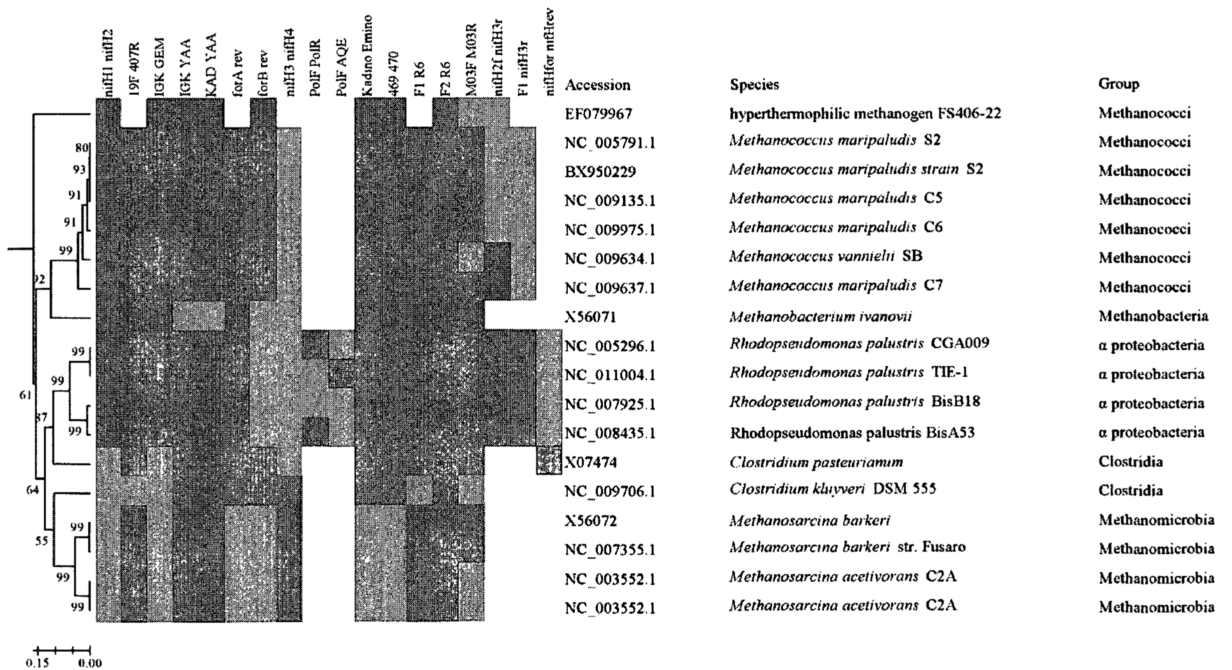
G



H



I



J