Minireview

Nitrogenase gene diversity and microbial community structure: a cross-system comparison

Jonathan P. Zehr,[1]* **Bethany D. Jenkins,**[1]
Steven M. Short[1] **and Grieg F. Steward**[2]
[1]*Department of Ocean Sciences, University of California, Santa Cruz, Santa Cruz, CA 95064, USA.*
[2]*School of Ocean and Earth Science and Technology, University of Hawaii, Manoa, Honolulu, HI 96822.*

Summary

Biological nitrogen fixation is an important source of fixed nitrogen for the biosphere. Microorganisms catalyse biological nitrogen fixation with the enzyme nitrogenase, which has been highly conserved through evolution. Cloning and sequencing of one of the nitrogenase structural genes, *nifH*, has provided a large, rapidly expanding database of sequences from diverse terrestrial and aquatic environments. Comparison of *nifH* phylogenies to ribosomal RNA phylogenies from cultivated microorganisms shows little conclusive evidence of lateral gene transfer. Sequence diversity far outstrips representation by cultivated representatives. The phylogeny of nitrogenase includes branches that represent phylotypic groupings based on ribosomal RNA phylogeny, but also includes paralogous clades including the alternative, non-molybdenum, non-vanadium containing nitrogenases. Only a few alternative or archaeal nitrogenase sequences have as yet been obtained from the environment. Extensive analysis of the distribution of *nifH* phylotypes among habitats indicates that there are characteristic patterns of nitrogen fixing microorganisms in termite guts, sediment and soil environments, estuaries and salt marshes, and oligotrophic oceans. The distribution of nitrogen-fixing microorganisms, although not entirely dictated by the nitrogen availability in the environment, is non-random and can be predicted on the basis of habitat characteristics. The ability to assay for gene expression and investigate genome arrangements provides

the promise of new tools for interrogating natural populations of diazotrophs. The broad analysis of nitrogenase genes provides a basis for developing molecular assays and bioinformatics approaches for the study of nitrogen fixation in the environment.

Introduction

Biological nitrogen (N_2) fixation, which catalyses the reduction of atmospheric N_2 gas to biologically available ammonium, is ecologically important as an input of fixed nitrogen (N) in many terrestrial and aquatic habitats (Vitousek and Howarth, 1991; Arp, 2000), and has been studied in a variety of ecosystems using multiple approaches to measure rates and characterize diazotroph populations. N_2 fixation is fundamentally important because it makes atmospheric dinitrogen (N_2) available to relieve ecosystem N limitation. Nitrogen limiting conditions should select for N_2 fixers, which can then drive ecosystems to limitation by other elements, such as phosphorus. If N limitation does occur, then there must be other factors that limit N_2 fixation (Vitousek, 1999), such as Fe, P or Mo availability (Howarth and Marino, 1988; Wu *et al.*, 2000). The balance between N_2 fixation and the reverse process, denitrification, can determine the net biologically available N for the biosphere (Arp, 2000). Thus, N_2 fixation interacts with other biogeochemical cycles to control the N status of the ecosystem.

Microbial communities are a fundamental component of ecosystems that play critical roles in metabolism of organic matter and in biogeochemical transformations of elements, including N_2 fixation (Atlas and Bartha, 1998; Madigan *et al.*, 2000). We have attained a new appreciation for the genetic diversity in natural microbial communities through the application of molecular biological techniques. Analysis of ribosomal RNA (rRNA) phylotypes has begun to uncover the pattern of global distribution of major groups of certain phylotypes such as the beta proteobacteria in freshwater (Methe *et al.*, 1998) and the SAR11 cluster in the open ocean (Mullins *et al.*, 1995), and the distribution of individual genotypes (Cho and Tiedje, 2000). These rRNA-based investigations led to the startling realization that we do not have cultivated representatives of the

numerically most abundant microorganisms in the environment (Pace, 1997), but they identified some important targets for renewed, more focused cultivation efforts (Bull *et al.*, 2000; Rappe *et al.*, 2002).

Although it has revolutionized our view of microbial diversity, rRNA analysis alone is insufficient for a complete understanding of microbial ecology as it does not provide definitive information on biological, physiological or ecological function. Nitrogenase genes, for example, are found throughout both the Archaea and Bacteria, but are sporadically distributed within clades in most cases (Young, 1992) such that phylogenetic affiliation can rarely be used to infer N_2 fixing capability.

Most microorganisms that perform biological N_2 fixation do so with an evolutionarily conserved nitrogenase protein complex (Howard and Rees, 1996). The high degree of similarity of protein sequence of nitrogenases among microorganisms suggests an early origin or lateral gene transfer among prokaryotic lineages (Postgate and Eady, 1988). As nitrogenase and 16S rRNA sequences accumulate in the databases, the high degree of sequence similarity supports early evolution of nitrogenase in an early ancestor (Postgate and Eady, 1988; Young, 1992). Current research on a number of genes (Doolittle, 1999) and comparative genomics, however, once again raises the speculation that nitrogenases could have been dispersed by lateral gene transfer mechanisms. Further detailed comparative phylogenetic analyses will ultimately help resolve the relative evolutionary histories of nitrogenases and rRNA genes. Regardless of the end result, nitrogenase gene sequences provide us, at present, with a practical means of classifying and identifying uncultivated diazotrophic microorganisms.

The nitrogenase enzyme is composed of two multisubunit metallo-proteins. Component I contains the active site for N_2 reduction, has a molecular weight of about 250 kDa and is composed of two heterodimers, encoded by the *nifD* and *nifK* genes. Component II (about 70 kDa) couples ATP hydrolysis to interprotein electron transfer and is composed of two identical subunits encoded by the *nifH* gene. Fe-S centres are present in both Component I and Component II proteins and are co-ordinated between the subunits. 'Conventional' nitrogenases contain Mo in the Fe-S centre bridging the subunits. 'Alternative' nitrogenases replace Mo with V (*vnfH*), and 'second alternative' nitrogenases replace Mo with Fe (*anfH*). The enzymes have somewhat different reaction kinetics and specificities (Burgess and Lowe, 1996; Eady, 1996). The nitrogenase reaction is energetically expensive (16 ATP and eight electrons per molecule reduced) and the enzyme *in vitro* is sensitive to inactivation by oxygen.

Recently, another unusual nitrogenase that is linked to carbon monoxide reductase activity has been described in *Streptomyces thermoautotrophicus* (Ribbe *et al.*, 1997).

Interestingly, the CO dehydrogenase of *Streptomyces* uses O_2 to generate O_2^- radicals that are used as electron carriers. This organism was isolated from an unusual CO-rich environment, and the distribution of this oxygen-insensitive enzyme in nature is not yet known.

In addition to the structural genes for the conventional nitrogenase, a number of other gene products are required for regulation and assembly. Thus, N_2 fixation is both metabolically expensive and can involve on the order of 20 kb of DNA or more to encode all of the proteins required for assembly and function. One might suppose that the presence of the *nif* genes would depend on selection for N_2 fixing capability, because the multiple genes involved in regulation and assembly occupy a fairly large piece of genomic real estate. Nitrogenase gene expression is highly regulated (Hoover, 2000), at levels ranging from transcription (Chen *et al.*, 1998) to post-translational protein modification (Kim *et al.*, 1999). Transcription of the *nifHDK* operon is a good marker for N_2 fixing conditions, as it is not constitutively expressed and is regulated in response to factors that control N_2 fixation.

Culture-independent methods have been used to investigate N_2 fixation in many different habitats by focusing on molecular sequences of the nitrogenase genes or using probes to ribosomal RNA phylotypes that are known to fix N_2 (Simonet *et al.*, 1991; Young, 1992; Ueda *et al.*, 1995a; Ohkuma *et al.*, 1996; Steppe *et al.*, 1996; Zehr *et al.*, 1998; Lovell *et al.*, 2000; Tan *et al.*, 2001; Hurek *et al.*, 2002). These studies have focused on habitats ranging from invertebrate guts, soils, plants, bioreactors, lakes, and rivers to the open ocean, and have uncovered a great diversity of nitrogenase sequences. The database for nitrogenase genes (specifically the *nifH* gene) has become one of the largest non-ribosomal gene datasets on uncultivated microorganisms.

The goals of this review are to examine the ecological and biological patterns of N_2-fixing microorganisms in a wide variety of habitats spanning the globe and to synthesize this information to provide the backbone for future studies of the diversity of N_2-fixing microorganisms in the Earth's biosphere. In this paper, we examine the broad patterns of N_2-fixing phyla across multiple environments as a basis for understanding the factors that determine the distribution of N_2 fixation potential, and more generally to examine the factors that determine the distribution of genes in genomes as a function of the selective forces in the environment.

Methods for assaying nitrogenase diversity

The study of nitrogenase diversity thus far has been largely based on the phylogenetic analysis of *nifH* (Zehr and McReynolds, 1989) and sometimes *nifD* (Ueda *et al.*, 1995b) gene sequences. *NifD* and *nifK* will ultimately be

useful as it is not clear that the phylogeny of *nifH, D* and *K* are always consistent (Dominic *et al.*, 2000). However, there are relatively few *nifD* and *K* sequences available, so the use of these genes at this time is limited in scope for phylogenetic analysis or phylogenetic comparisons. These genes, when developed as phylogenetic markers, promise to provide more resolution among closely related strains and better differentiate different *nif* gene family members such as the V nitrogenases.

Samples are usually collected for DNA, or more recently mRNA (Ohkuma *et al.*, 1999; Zani *et al.*, 2000; Tan *et al.*, 2001) and various procedures are used to obtain pure nucleic acid extracts that do not inhibit reverse transcription (RT) or the polymerase chain reaction (PCR). A number of *nifH* PCR primer sets have been designed, although most of them target the same or overlapping sites and differ primarily in their use of degenerate oligonucleotides or modified nucleotides (Poly *et al.*, 2001a). In some cases, the primers have been designed on the basis of specific anticipated phylogenetic groups (for example ignoring cyanobacteria sequences in the marsh studies). The nucleic acid extracts are amplified in RT-PCR or PCR reactions resulting in mixed pools of amplicons that reflect the diversity of nitrogenase gene sequences in the environment. Although these amplicons can be analysed by a number of fingerprinting techniques including DGGE (Piceno and Lovell, 2000a), terminal restriction fragment length polymorphism (Ohkuma *et al.*, 1999) or DNA microarrays, conventional cloning and sequencing yields a library of molecular sequences that can be used for detailed phylogenetic analysis. The use of different primers in different studies may result in biases, and primer bias in general has not been evaluated quantitatively in most studies (Zehr and Capone, 1996). The effect of primer bias should be kept in mind in the analyses presented here, although in general the similarities between studies are more remarkable than the differences. Most studies have attempted to design 'universal' primers, but not all. The sequences recovered from different habitats may reflect differing PCR conditions (e.g. numbers of cycles) used in different studies, as well.

The *nifH* sequence database is rapidly expanding and is currently composed of over 1500 sequences, most of which have been obtained from environmental samples. This database provides a resource for developing RT-PCR, DNA array and quantitative PCR approaches, providing that the distribution of sequences is representative of the environment. One of the objectives of this review is to assess the distribution of *nifH* phylotypes across systems, which will help to determine where cultivation efforts should be focused and aid in the identification and development of probes for the environment.

The large size of the database and the rapidity with which it is growing has necessitated the parallel develop-

ment of bioinformatics approaches to provide a consistent foundation for comparing the results of phylogenetic analyses. To ensure consistent alignment of *nifH* sequences we are using a profile hidden Markov model (HMM) built iteratively with HMMER 2.2 (*http://hmmer.wustl.edu*; now also available in the GCG software package, Wisconsin Package Version 10.3, Accelrys, San Diego, CA). To build the model we downloaded the fer4_NifHPfam seed alignment (*http://pfam.wustl.edu*) and made minor manual adjustments. An initial model built on the modified seed alignment was then used to align a selection of 50 additional protein sequences chosen to represent a wide diversity of known *nifH* sequence types. The resulting alignment was evaluated by eye, subjected to minor manual corrections and used to build a second profile HMM. This second-iteration HMM was then used to align protein translations of the full *nifH* database which was composed of sequences available in GenBank in addition to sequences we recently obtained from several aquatic environments that will be published elsewhere (B. Jenkins, G. Steward, E. Omoregie, L. Crumbliss and J. Zehr, unpub. data, GenBank Accession numbers AY221756–AY221832, AY223907–AY224045, AY221811–AY221832, AY244717–AY244740 and AY232333–AY232412). The aligned protein sequences were imported into ARB (Strunk and Ludwig, 1993) and used to align the corresponding DNA sequences.

Major *nifH* clusters

Nitrogenase genes cluster into four basic groups, previously designated as Clusters I-IV (Chien and Zinder, 1996)(Fig. 1A). The composition of these major clusters consist of the 'conventional' Mo-containing *nifH* and some *vnfH* (Cluster I), the 'second alternative' non-Mo, non-V containing *anfH* as well as nitrogenases from some Archaea (Cluster II), *nifH* sequences from a diverse group of distantly related microorganisms many of which are strict anaerobes (e.g. clostridia, sulphate reducers; Cluster III), and a divergent loosely coherent group of *nif*-like sequences from Archaea and distantly related chlorophyllide reductase genes (Cluster IV). It is not clear whether the *vnfH* genes within Cluster I (Table 1) can be distinguished from conventional *nifH*. However, the *vnfH* genes of *Azotobacter* species cluster together (Cluster 1 N, Fig. 1), and as more *vnfH* genes are identified and sequenced it may be possible to resolve *nifH* from *vnfH* sequences.

Many microorganisms have multiple copies of nitrogenase genes or homologues of nitrogenase genes (Cluster IV). *Clostridium pasteurianum* was shown to have a *nifH* gene family (Wang *et al.*, 1988), and has sequences in Clusters II and III (Table 1). *Azotobacter* was the first genus to be shown to have both alternative nitrogenases,
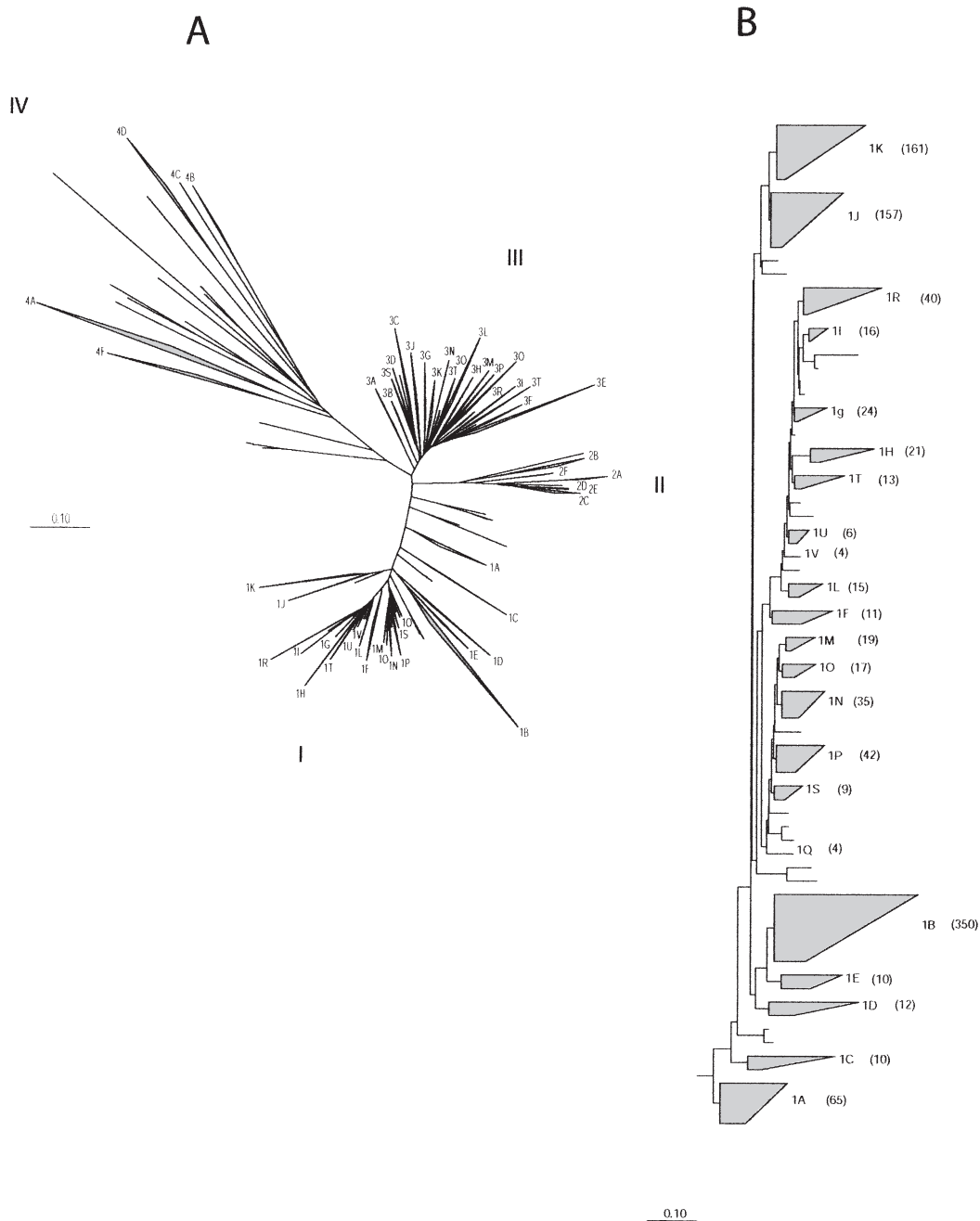
Fig. 1. Phylogenetic distribution of *nifH* genes based on neighbour-joining analysis of partial amino acid sequences.
A. The four major clusters of *nifH* and *nifH*-like genes. Clusters I-III include functional nitrogenases.
B. Phylogenetic distribution of sequences in the current *nifH* database in Cluster I. Numbers in brackets indicate number of sequences in each cluster.

vnfH and *anfH* (Bishop *et al.*, 1985). A number of other examples of microorganisms with multiple copies of nitrogenase genes have been uncovered. Some cyanobacteria have multiple copies of nitrogenase, including a *vnfH* as well as a second distinct copy of *nifH* (Thiel, 1993; Thiel *et al.*, 1995). Sulphate reducers contain multiple homologues of *nifH* (Braun *et al.*, 1999) in which group in Clus-

ters II and III. Archaea often have two *nifH* homologues. In some cases, the two copies are in Cluster IV and Cluster III, and in other cases, the homologues are in Cluster II and Cluster III (Table 1). The placement of the multiple copies in the different phylogenetic clusters might be a phylogenetic diagnostic that could be exploited for the development of probes and primers.

Table 1. Phylogenetic distribution of *nifH* genes among cultivated representatives. Major phylogenetic groups based on rRNA and subclusters based on *nifH* are shown.

Cluster	Group	Subcluster	Genera	Cluster	Group	Subcluster	Genera
I	Alpha	1J	Azospirillum				Synechocystis (marine)
			Gluconacetobacter				Tolypothrix
			Mesorhizobium				Trichodesmium
			Rhodobacter				Xenococcus
			Rhodospirillum	I	Firmicutes	1D	Frankia
			Rhizobium			1E	Paenibacillus
			Sinorhizobium	II	Alpha	2C	Rhodobacter
		1K	Beijerinckia		Delta	2E	Desulfobacter
			Methylocella		Gamma	2C	Azotobacter
			Methylosinus		Firmicutes	2D	Paenibacillus
			Methylocystis			2E	Clostridium
			Rhizobium		Spirochaetes	2C	Spirochaeta
			Xanthobacter				Treponema
	Beta	1J	Burkholderia		Archaea	2B	Methanobrevibacter
		1K	Burkholderia				Methanococcus
			Herbaspirillum				Methanothermobacter
		1P	Azoarcus			2F	Methanosarcina
		1U	Alcaligenes	III	Delta	3B	Desulfobacter
	Epsilon	1F	Arcobacter			3E	Desulfomicrobium
	Gamma	1H	Vibrio				Desulfovibrio
		1K	Acidothiobacillus			3J	Desulfotomaculum
		1I	Klebsiella			3L	Desulfonema
			Vibrio			3P	Desulfovibrio
		1M	Marichromatium			3T	Desulfovibrio
			Methylomonas		Firmicutes	3A	Clostridium
		1N	Azotobacter (*vnfH*)			3C	Clostridium
			Klebsiella			3D	Acetobacterium
		1P	Methylobacter				Clostridium
		1T	Azomonas			3J	Desulfosporosinus
			Azotobacter		Spirochaetes	3C	Spirochaeta
		1U	Pseudomonas				Treponema
	Cyanobacteria	IB	Anabaena			3K	Spirochaeta
			Chlorogloeopsis			3L	Spirochaeta
			Calothrix				Treponema
			Cyanothece			3S	Spirochaeta
			Dermacarpa		Archaea	3C	Methanosarcina
			Fischerella		Green sulphur	3L	Chlorobium
			Gloeothece				Pelodictyon
			Lyngbya	IV	Spirochaetes	4A	Treponema
			Myxosarcina			4C	Treponema
			Nostoc		Archaea	4	Methanobrevibacter
			Oscillatoria				Methanocaldococcus
			Phormidium				Methanococcus
			Plectonema				Methanopyrus
			Pseudanabaena				Methanosarcina
			Scytonema				Methanothermobacter
			Symploca			4A	Methanosarcina
			Synechococcus (Cyanothece)			4D	Methanosarcina

NifH *phylogeny of cultivated diazotrophs*

nifH gene sequences from cultivated microorganisms, and their cluster affiliation are shown in Table 1. The proteobacteria *nifH* sequences form a number of distinct clusters (Fig. 1B and Fig. 2) that correspond approximately, but not perfectly, with ribosomal RNA phylogeny (Table 1, Fig. 1B). The gamma and alpha proteobacterial clusters are generally well-defined (Fig. 1B, Table 1), although there are a few cultivated microorganisms that do not consistently cluster within the appropriate group.

Cultivation techniques, although biasing for viable, cultivable cells, have provided crucial landmarks to ground-truth sequences obtained by amplification (Steppe *et al.*, 1996; Bagwell *et al.*, 1998; Egener *et al.*, 1999; Olson *et al.*, 1999; Lilburn *et al.*, 2001; Steppe and Paerl, 2002). For example, the *nif* sequences of spirochaetes isolated from termites were very similar to sequences obtained by direct amplification (Ohkuma *et al.*, 1996; Lilburn *et al.*, 2001). Isolates have also been useful for developing *in situ* hybridization assays for enumerating and visualizing N$_2$-fixing cells *in situ* (Hurek *et al.*, 1997; Egener *et al.*, 1998).

Archaeal nitrogenases have thus far only been found in methanogens, and though the presence of nitrogenase in other Archaeal lineages is likely, it has yet to be documented. Archaeal nitrogenases include homologues of
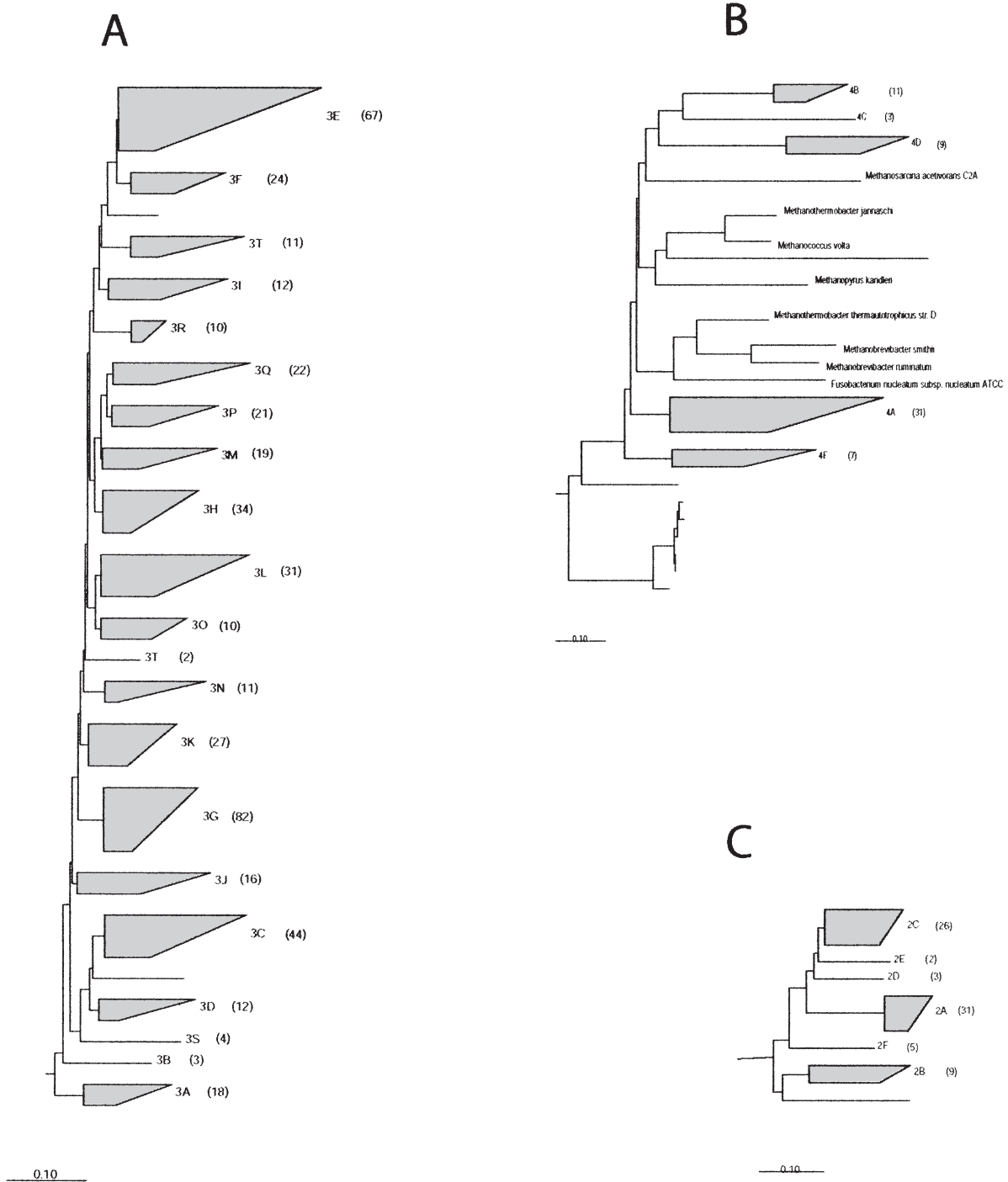
Fig. 2. Phylogenetic distribution of *nifH* and *nifH*-like genes. Clusters III (A), IV (B) and II (C). Numbers in brackets indicate number of sequences in each cluster.

both functional nitrogenases and *nif*-like genes of unknown function (Chien *et al.*, 2000).

Methylotrophs of both types I and II have now been shown to have *nifH* and to be capable of N$_2$ fixation (Auman *et al.*, 2001) (Table 1, Groups 1K, 1M, 1P). Interestingly,

nifH sequences from strains isolated from the environment are very similar (>94% identical at the amino acid level) to sequences amplified from a number of habitats, implying that methylotrophs may be found to be important in N as well as carbon cycling (Auman *et al.*, 2001).

Cyanobacteria are a morphologically and physiologically diverse group within the Bacteria. The cyanobacteria *nifH* genes cluster together (Zehr *et al.*, 1997) (Fig. 1B, Cluster 1B) although sequences from some unicellular and filamentous non-heterocystous cyanobacteria form deep branches. Heterocyst-forming cyanobacterial *nifH* form a tight cluster within the cyanobacterial group. Some cyanobacteria have a nitrogenase that is expressed in vegetative cells of filamentous heterocyst-forming species (Thiel *et al.*, 1995). Vanadium nitrogenases have also been reported in cyanobacteria and the *Anabaena variabilis vnfH* is found within the heterocyst-forming cyanobacterial *nifH* clade (Thiel, 1993). As most nitrogenases from cyanobacteria have not yet been characterized by mutant or deletion analysis, it is not yet possible to determine whether the vegetative, heterocyst or *vnfH* nitrogenases in cyanobacteria (Cluster 1B) can be resolved by *nifH* phylogeny. Thus, the cyanobacterial *nifH* cluster (Fig. 1B, cluster 1B) contains *nifH* genes expressed in vegetative cells and heterocysts, as well as the cyanobacterial *vnfH* genes.

The cyanobacteria gene cluster can intermingle with *Frankia nifH* sequences of the firmicutes (Hirsch *et al.*, 1995), although as more sequences have been added to the database, the resolution of these groups has improved.

Paenibacillus species are Gram-positive bacteria which are typically found in rhizosphere zones. Analysis of *nifH* genes in *Paenibacillus* species with a nested PCR approach showed that these organisms also contain multiple copies, with one homologue in Cluster II (Table 1)(Rosado *et al.*, 1998).

The sequences from a number of cultivated diazotrophs are in Cluster III. These sequences include those from Gram-positive microorganisms, delta proteobacteria, green sulphur bacteria and Archaea. The organisms represented by *nifH* within this cluster are mostly, if not all, strict anaerobes. The known organisms in this cluster are *Clostridium*, *Desulfovibrio (*and other sulphate reducing genera), *Chlorobium*, and the Archaea *Methanosarcina barkeri*. Although they cluster together, Cluster III is characterized by deep bifurcations and long branch lengths, and the distances between sequences are large relative to distances within Cluster I (Fig. 2). The clustering of these sequences is not as aberrant with ribosomal phylogeny as it might first appear, as phylogenetic trees based on partial 16S ribosomal sequences from these microorganisms groups them together as well (Fig. 3).

Sulphate reducers are known to be N_2 fixers and have nitrogenase genes (Kent *et al.*, 1989) that are in Cluster III. Experimental data implicated them in N_2 fixation in estuarine sediments (Dicker and Smith, 1980). However, it is curious that sulphate reducers fix N_2 in many cases, because they are active in highly reduced sediments where decomposition processes release ammonium.

Steppe and Paerl (2002) investigated the involvement of sulphate reducers in nitrogen fixation in marine sediments and provided evidence for *nifH* gene expression.

Cluster II and III *nifH* sequences from spirochaetes have been recently reported (Lilburn *et al.*, 2001) (Table 1). *nifH* genes from termites guts that are closely related to spirochaete *nifH* genes (Lilburn *et al.*, 2001) were shown to be transcribed using RT-PCR (Noda *et al.*, 1999; Ohkuma *et al.*, 1999).

Representative sequences for several major lineages are still not represented on the phylogenetic tree. For example, there are no nitrogenases from bacterial thermophiles (except from the green sulphur bacteria), yet it seems likely, given the dispersion of *nifH* through the prokaryotes, that they will be present in this group. Although it is possible that some of these lineages do not have N_2-fixing representatives, it is probably still true that we simply have not yet provided cultures with the right conditions to observe N_2-fixation.

Subclusters of nifH *sequences from cultivated and uncultivated microorganisms*

Within the four major clusters, sequences group within subclusters that have been tentatively defined on the basis of amino acid sequences for the purposes of this cross-system comparison (Table 1, Figs 1B and 2). The sequences from cultivated microorganisms help to classify sequence types obtained from the environment and evaluate the topology of the tree with respect to the evolution of nitrogenase. The conclusion of many surveys of diazotroph diversity is that assemblages are diverse, and that the phylotypes obtained directly from the environment are not closely related to sequences from previously characterized strains and environments.

Lateral gene transfer

Recently, particularly in light of genomic information, there has been discussion of the potential for lateral gene transfer among microorganisms (Doolittle, 1999; Nesbo *et al.*, 2001). Dissimilatory sulphite reductase (DSR), the gene encoding the protein catalysing the last step in respiratory sulphate reduction, is one candidate for lateral gene transfer based on comparison of DSR and 16S rRNA phylogenies, and between DSR subunits (Klein *et al.*, 2001). Nitrogenase is another functional gene that could easily be subject to lateral gene transfer, and in fact, exists on plasmids in some rhizobia (Barbour *et al.*, 1992). However, most of the nitrogenase genes that have been characterized by genomic analysis are not plasmid-borne, despite the fact that diazotroph assemblages do contain plasmids (Beeson *et al.*, 2002). It is clear that true evolutionary relationships cannot necessarily be inferred from
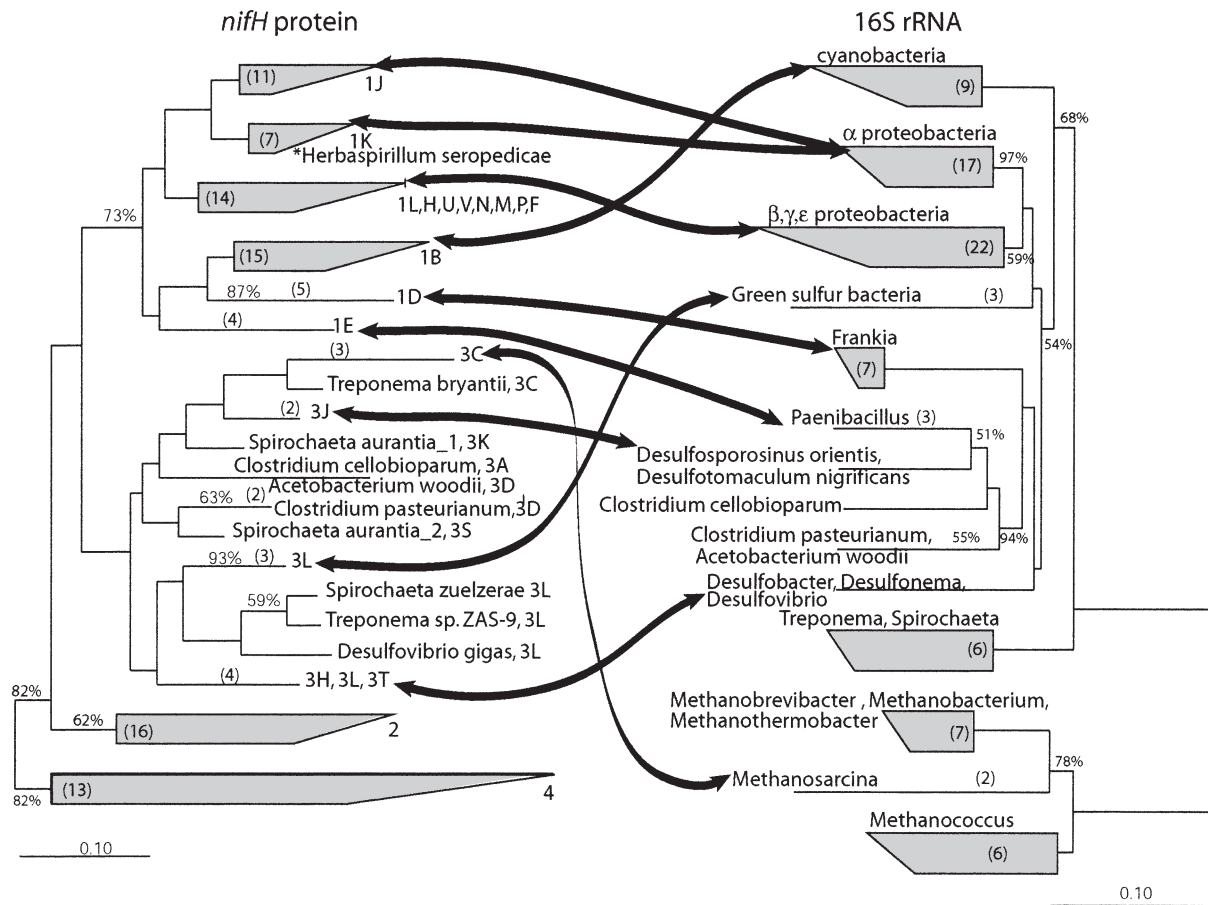
Fig. 3. Comparison of *nifH* (A) and 16S rRNA (B) phylogeny for cultivated microorganisms. *nifH* and 16S rRNA sequences were aligned and analysed by neighbour-joining using ARB. Analysis was bootstrapped and bootstrap values greater than 50% are indicated at respective nodes. Numbers in brackets indicate number of sequences in each cluster. *nifH* cluster names refer to cluster designations shown in Figs 1 and 2.

either single gene trees or by integrating the information from multiple gene trees (Cavalier-Smith, 2002), but here we compare the phylogeny of *nifH* to that of 16S rRNA genes for cultivated microorganisms for which both sequences are available, in order to examine the evidence for lateral gene transfer of *nifH*.

The list of cultivated organisms for which there are *nifH* gene sequences in GenBank was used to extract aligned 16S rRNA sequences for the same or closely related representatives from the Ribosomal Database Project (Maidak *et al.*, 2001). The aligned sequences were used to generate a phylogenetic tree using ARB, so that the ribosomal RNA phylogeny could be directly compared to *nifH* phylogeny.

Trees constructed for *nifH* and 16S genes based on the same lineages represented by both genes generate remarkably consistent trees (Fig. 3). It is difficult to assess transfers in Archaea, because Clusters II and IV include many of the Archaeal *nifH* genes, but are paralogous to the *nifH* genes in Clusters I and III. In fact, Cluster III could

itself be a paralogue arising from a duplication of *nifH* early in evolution (Hirsch *et al.*, 1995). Nonetheless, the topology between the two trees is largely consistent (Fig. 3). The gamma group including *Vibrio* and *Klebsiella* is congruent between trees, as is the cyanobacterial group. *Frankia* and *Paenibacillus* form distinct groups in both trees (Fig. 3), although *Paenibacillus* clusters with a larger group including *Clostridium* in the 16S rRNA tree, but does not on the basis of *nifH*.

Rhizobium, Mesorhizobium, Azospirillum Rhodobacter and *Methylosinus* cluster together on both trees as a group of alpha proteobacteria. *Herbaspirillum* is perhaps the best example of aberrant phylogeny, grouping with *Alcaligenes* and *Azoarcus* in the ribosomal RNA tree. The Cluster III sequences in the *nifH* tree include sequences from *Clostridium, Chlorobium*, some spirochaetes, and a number of sulphate respirers (e.g. *Desulfonema* and *Desulfobacter*). These sequences also form distinct clusters in the 16S tree when only sequences from N₂ fixing microorganisms are analysed (Fig. 3). Thus, although this clus-

Table 2. Average and standard deviation of distances within representative phylogenetic clusters based on *nifH* amino acid (AA), DNA and 16S rRNA sequences. Genetic distances calculated in ARB.

	Average			STD		
	Nif AA	Nif DNA	16S	Nif AA	Nif DNA	16S
alpha	0.127787	0.208548	0.129397	0.083251	0.07886	0.056603
gamma	0.129437	0.26234	0.13127	0.084239	0.082884	0.052578
cyanobacteria	0.105378	0.228298	0.117204	0.047359	0.062929	0.031335
Frankia	0.153447	0.228292	0.036254	0.153778	0.146156	0.018789
Archaea (methanogens)	0.417256	0.603736	0.184933	0.130475	0.603736	0.126849

ter of sequences initially seemed to contradict rRNA phylogeny (Zehr *et al.*, 1995), it is perhaps largely a result of the sequences included in phylogenetic analysis.

Genetic distances among sequences of major clusters are greater on the basis of *nifH* than 16S rRNA sequences (Table 2). The regions used for *nifH* PCR amplification (included in this analysis) include the more divergent regions of the *nifH* gene. The nitrogenase gene is likely not to be as conserved as rRNA, and this data shows the difference in rates of evolution between genes.

This analysis is based on a relatively short (~350 nt), but phylogenetically informative, region of the *nifH* gene, and the resolution of the issue of lateral gene transfer will depend on more extensive genomic analyses, or possibly analyses based on the molecular sequences of the other nitrogenase structural genes. Nonetheless, there is not strong support or evidence for lateral gene transfer, by plasmid transfer or other mechanisms, and if such transfers occurred, they must have occurred early in evolution. Finally, we are aware of at least two probable gene duplications, giving rise to the alternative nitrogenases. It is possible that other duplications exist within the *nifH* phylogenetic tree, that could lead to false conclusions about phylogeny based on paralogous comparisons. *NifD* and *nifK* analyses may help in this regard, as well as genomic context information (Young, 1992).

Phylogenetic distribution within habitats

Cross-system comparison of nifH *diversity*

The question of how diazotrophs are distributed across ecosystems and habitats can be investigated now that there is an extensive dataset of *nifH* genes. The diversity of *nifH* genes as a function of habitat is at the heart of biocomplexity, and encompasses the issue of how genetic redundancy relates to ecosystem function. This issue is also central to understanding how organisms and genes are selected in habitats, and how selection versus other factors determine the distribution and diversity of genotypes. Table 3 depicts the distribution of *nifH* genes obtained from different habitats within subclusters of *nifH* genes that contain cultivated representatives (Table 1).

Table 4 shows phylogenetic clusters that are currently composed of only environmental sequences. The vertical distribution of shaded areas shows how subcluster phylotypes are distributed in relation to the phylogeny of *nifH* (Figs 1 and 2).

Sequences have been obtained that group in Cluster I from many environments, consistent with the widespread distribution of proteobacteria (purple bacteria) and cyanobacteria. Cluster III is particularly interesting as multiple divergent sequences within Cluster III have been retrieved from environments that may be anaerobic at times or may contain anaerobic or microaerophilic microsites. These environments include estuaries, a hypersaline lake, sediments, microbial mats, termite guts and planktonic crustaceans (Table 3).

Interestingly, relatively few Cluster II alternative nitrogenase gene sequences have been obtained from the environment (Table 3). Primers directed at *anfG*, which is diagnostic for alternative nitrogenases, have demonstrated that alternative nitrogenase containing microorganisms are present in a few environments (Loveless and Bishop, 1999). Few sequences of this group have been obtained from amplification using *nifH* primers (Table 3), indicating their probable low relative abundance (on a gene level).

Very few sequences have been obtained from Cluster IV (Table 3). These sequences were amplified from termites, and could be homologues from spirochaetes or methanogens, both typical inhabitants of the termite gut.

It is clear that diazotroph diversity is not distributed equally among habitats, and that there are differences in distribution of specific types and the number of types across habitats (Table 3). There are many clusters for which we do not yet have cultivated isolates (Table 4), and some of these clusters are found in multiple environments (e.g. 1 A, 1G, 3H, 3I; Table 4). Some habitats have less *nifH* sequence diversity, although this may be partially due to discrepancies in the number of sequences obtained in each study. Marshes, mats, termites and the aquatic water column have received the most sequencing efforts. However, a number of distinct patterns emerge independent of the numbers of sequences obtained from a given environment.

Table 3. Subclusters (see Figs 1 and 2) of *nifH* phylotypes obtained from the environment that contain cultivated N$_2$ fixing microorganisms (data available as of July, 2002). Grey boxes indicate clusters with sequences from environments indicated in columns. Numbers in brackets indicate number of sequences in each group or cluster.

Subcluster (# of sequences)	Antarctic (14)	Lake (82)	Open Ocean (181)	Estuary (206)	Marsh (244)	Sea Grass (69)	Microbial Mat (352)	Stromatolites (16)	Sediment (16)	Forest Soil (29)	Rice Root (26)	Marine Invertebrate (18)	Termite (165)
1B (350)	■	■	■	■			■	■					
1D (12)						■							
1E (10)									■				
1F (11)		■		■								■	■
1H (21)	■												
1J (157)	■	■	■	■			■			■			
1K (161)		■	■	■	■	■		■					
1L (15)			■			■							
1M (19)		■											
1N (35)			■										
1P (42)		■				■	■		■		■		■
1T (13)		■		■		■			■				
1U (6)			■										
2C (26)					■								■
3A (18)													■
3B (3)					■								
3C (44)													■
3D (12)												■	
3E (67)	■	■		■			■	■					
3J (16)				■			■						
3K (27)		■				■					■	■	
3L (31)		■						■			■	■	
3P (21)		■			■		■						
3S (4)		■											
4A (31)													■
4C (3)													■
4D (9)													■

Oligotrophic oceans

In oligotrophic aquatic habitats, *nifH* genes do not appear to be as abundant as in productive lakes or coastal waters or estuaries, or sediments. Interestingly, different groups of *nifH* genes have been obtained from the open ocean than from lakewaters, such as from Lake Michigan or Lake George (Tables 3 and 4). Relatively few different major groups have been found in the plankton (Tables 3 and 4), but interestingly new species of diazotrophs have been discovered on the basis of the *nifH* gene and transcripts (Zehr *et al.*, 2001). Cyanobacteria *nifH* genes observed have included the well-known marine diazotroph *Trichodesmium*, but new groups of genes from unicellular cyanobacteria, some of which are now in culture (Zehr *et al.*, 2001), have routinely been detected in Pacific Ocean waters at station ALOHA. Other groups include *nifH* genes from yet undescribed heterocyst-forming cyanobacteria and gamma proteobacteria.

Lakes, rivers and estuaries

N$_2$ fixing cyanobacteria species are often involved in algal blooms in lakes and rivers (Paerl, 1988). Although the amplification of *nifH* can be useful in identifying N$_2$ fixing species (Zehr and McReynolds, 1989; Olson *et al.*, 1999; Steppe *et al.*, 2001; Zehr *et al.*, 2001), or identifying species expressing nitrogenase (Zani *et al.*, 2000; 2001), it is rather short for a useful phylogenetic marker in closely related strains (Zehr *et al.*, 1997; Tamas *et al.*, 2000; Dyble *et al.*, 2002). Nonetheless, Dyble *et al.* (2002) found that amplified *nifH* sequences from cyanobacterial isolates reflected their geographic origin. In depth profiles in Lake Michigan, cyanobacteria sequences were prominent, although sequences from bacterial groups were also found. Similarly both cyanobacterial and bacterial *nifH* genes were found in meso-oligotrophic Lake George, New York (Zani *et al.*, 2000).

An interesting, contrasting aquatic environment is Mono Lake, a hypersaline, currently meromictic lake in the eastern Sierra Nevada mountains of Northern California (Jellison and Melack, 2001). This lake is believed to be N limited, although attempts to detect N$_2$ fixation in the water column have been unsuccessful. Diversity of diazotrophs is high, and there is a substantial difference between the

Subcluster (# of sequences)	Antarctic (14)	Lake (82)	Open Ocean (180)	Estuary (206)	Marsh (244)	Sea Grass (69)	Microbial Mat (352)	Stromatolites (16)	Sediment (16)	Forest Soil (29)	Rice Root (26)	Marine Invertebrate (18)	Termite (165)
1A (65)		■		■	■						■		
1C (10)		■			■				■		■		
1G (24)		■			■							■	
1I (16)			■										
1O (17)				■	■				■				
1Q (4)	■			■									
1R (40)				■									
1S (9)		■		■		■							
1V (4)						■							
2A (31)				■							■		■
3F (24)							■						
3G (82)					■	■	■						
3H (34)				■	■		■					■	
3I (12)				■	■						■		
3M (19)		■				■	■						
3N (11)	■					■						■	
3O (10)	■						■		■				
3Q (22)					■		■	■					
3R (10)							■	■					
3T (11)							■						
4B (11)													■
4F (7)													■

Table 4. Novel subclusters (see Figs 1 and 2) of *nifH* phylotypes obtained from the environment that are defined only by sequences obtained from the environment and do not contain sequences from cultivated microorganisms. Grey boxes indicate clusters with sequences from environments indicated in columns. Numbers in brackets indicate number of sequences in each group or cluster.

communities of the mixolimnion and monimolimnion (G. Steward, unpub. data). Composition of these communities is substantially different than that found in other aquatic environments. For example, about two-thirds of the sequences obtained from Mono lake fall in Cluster III, a cluster that contains only two sequences from freshwater lakes and none from the open ocean. Several sequences group closely to *Desulfovibrio* sequences and one appears to be closely related to Spirochaetes, which is of interest because there were 16S sequences recovered from Mono Lake that group with Spirochaetes (Humayoun *et al.*, 2003). Of the Cluster III sequences recovered, some cluster within groups that contain only sequences recovered from the environment (3M,N,O,Q), specifically estuaries, salt marshes, sea grasses and microbial mats.

Estuaries have diverse groups of microorganisms including those that are similar to the distribution in marshes and mats. Two estuarine habitats where approximately the same number of *nifH* sequences have been analysed are the Chesapeake Bay (B. Jenkins, unpub. data) and the Neuse River (Affourtit *et al.*, 2001). Both of these environments contain large numbers of Cluster 1J and 1K alpha proteobacteria in regions where salinity is low. Both estuaries also contain other Cluster I beta and gamma proteobacteria sequences (1P,1U) and the Neuse River contains a small number of epsilon proteobacteria

sequences (cluster 1F). Unlike the Chesapeake Bay, where no cyanobacterial sequences (IB) have been recovered to date, about one-quarter of the Neuse River *nifH* sequences are from cyanobacteria. Both the Chesapeake Bay and the Neuse River have sequences in Cluster III (E,J,K,L,P). Sequences from Chesapeake Bay and Neuse River sediments are also found in Cluster III, which suggests that some of these Cluster III sequences may be derived from sediment mixing (Affourtit *et al.*, 2001; Burns *et al.*, 2002).

Mats and sediments

Mats and stromatolites are representative of microbial communities on the early Earth (Des Marais and Walter, 1999). These complex assemblages include heterotrophic and phototrophic microorganisms, and diazotrophs are found in both physiological types (Zehr *et al.*, 2001; Steppe and Paerl 2002). We have done extensive sequence analysis of *nifH* genes in mats in the hypersaline lagoons of Guerrero Negro (E. Omoregie and B. Bebout, unpubl. data). These sequences include Cluster I proteobacteria and cyanobacterial sequences as well as many sequences in Cluster III. Sequences from both Cluster I and III are expressed in the mats (E. Omoregie and B. Bebout, unpubl. data).

Sediment and salt marsh environments have diverse assemblages of *nifH* sequences, although there appear to be selection for different phylotypes between habitat types. Salt marsh ecosystems are important wetland environments at the land–sea interface, characterized by salt tolerant grass species. N_2 fixing microbial assemblages are diverse in these environments (Piceno and Lovell, 2000a,b), with a variety of phylogenetic and physiological types represented (Bagwell *et al.*, 1998). *nifH* genes cluster in Cluster I alpha and gamma proteobacteria groups as well as in Cluster III (Lovell *et al.*, 2001). Interestingly, diazotroph assemblages appear to be relatively stable even in the face of N enrichments (Piceno and Lovell, 2000b) and environmental variability (Bagwell and Lovell, 2000), as is N_2 fixation as assayed by acetylene reduction (Bagwell and Lovell, 2000; Piceno and Lovell, 2000a). Despite species stability, there are differences in species assemblages for different grasses (Bergholz *et al.*, 2001), above-ground versus below-ground (Lovell *et al.*, 2001) and for dead salt marsh wrack (Lovell *et al.*, 2000).

Soils

Interestingly, soils have fewer phylogenetic clusters represented than some aquatic environments (Tables 3 and 4). Using DGGE, it was found that the composition of the diazotroph community varied as a function of soil type on a range of scales (Poly *et al.*, 2001b). Phylotypes recognized by RFLP of amplified fragments indicated that diazotroph species in forest litter shift in clearcut communities (Shaffer *et al.*, 2000).

Rhizosphere soils contain bacteria of a number of anaerobic and aerobic bacteria that have been found in other environments including rice root rhizospheres and aquatic environments (Hamelin *et al.*, 2002). *Azoarcus* is an interesting genus composed of disparate lineages, some of which cluster in different proteobacterial clusters on the basis of *nifH* (Hurek *et al.*, 1997). *Azoarcus* sp. BH72 was shown to express nitrogenase in rice root rhizospheres, using a green fluorescent protein reporter (Egener *et al.*, 1998) and also inside rice roots (Egener *et al.*, 1999). Analysis of expression of *nifH* phylotypes in conjunction with natural abundance of ^{15}N showed that *Azoarcus* in Kallar grass (a productive grass which requires little N fertilization) was contributing to N_2 fixation.

Invertebrate-associated

Termite guts represent an extremely N-limited environment, as some termites can feed solely on cellulose and must obtain N elsewhere (Ohkuma *et al.*, 1999). Diverse *nifH* genes were recovered from termite guts using *nifH* PCR primers, including sequences that could be derived from methanogens (Ohkuma *et al.*, 1996). Sequences were recovered from Clusters II, III and IV. Undoubtedly some of these sequences represent multiple copies, because in many organisms, there are *nifH* homologues in groups IV and III, or in II and III such as in *Methanosarcina* (Chien and Zinder, 1996). Lilburn *et al.*, (2001) isolated spirochaetes of the genera *Treponema* and *Spirochaeta* from termites, showed that some of the isolates were capable of N_2 fixation, and found that the *nifH* genes clustered with previously reported *nifH* sequences from termite guts, as well as marine copepods.

It was shown that second alternative nitrogenases were expressed in termite guts on the basis of RT-PCR. Unexpectedly, the *anf* genes were not regulated by Mo as anticipated, as Mo usually represses the transcription of alternative nitrogenases (Noda *et al.*, 1999). Using a primer pair that bracketed *nifH* and *nifK*, PCR showed that the alternative nitrogenase in the termite guts contained *anfH*, two ORFs, and *anfDGK*, an arrangement consistent with Archaeal Cluster II nitrogenases. Nitrogenases were also found that are very similar to sequences found in other environments that cluster within Cluster I gamma proteobacteria sequences.

Factors limiting nitrogenase expression and environmental selection

N_2 is a potential source of N that should ultimately alleviate N limitation in the environment, unless other factors constrain N_2 fixation (Redfield, 1958; Howarth and Marino, 1988; Vitousek and Howarth, 1991; Vitousek, 1999). It is usually assumed that genes are ultimately not retained by microorganisms, unless they are functional and thus, are selected for in the environment. In the case of nitrogenase, this would seem to be particularly true because N_2 fixation can involve around 20 genes, even in free-living microorganisms (Fig. 4). It would seem likely that the presence and diversity of nitrogenase genes might reflect selection due to the extent of N limitation in the environment. One of the interesting patterns seen in Table 3 is that the diversity of nitrogenase genes is not directly linked to the degree of N limitation. The open ocean studies were performed in oligotrophic subtropical gyres where there are almost undetectable levels of fixed inorganic N (Karl and Michaels, 1996), and yet the diversity of *nifH* subclusters is less than in estuarine habitats such as the Chesapeake Bay and the Neuse River. Although the Neuse River is N limited in the outer reaches, upstream regions receive high N loading from agricultural activities (Steppe *et al.*, 2001). These data indicate that gene distributions in natural assemblages are a function of not only environmental selection but also physical-chemical factors including physical transport of cells between and among habitats. Microbial assemblages in estuaries may reflect an important allochthonous component, which has impli-
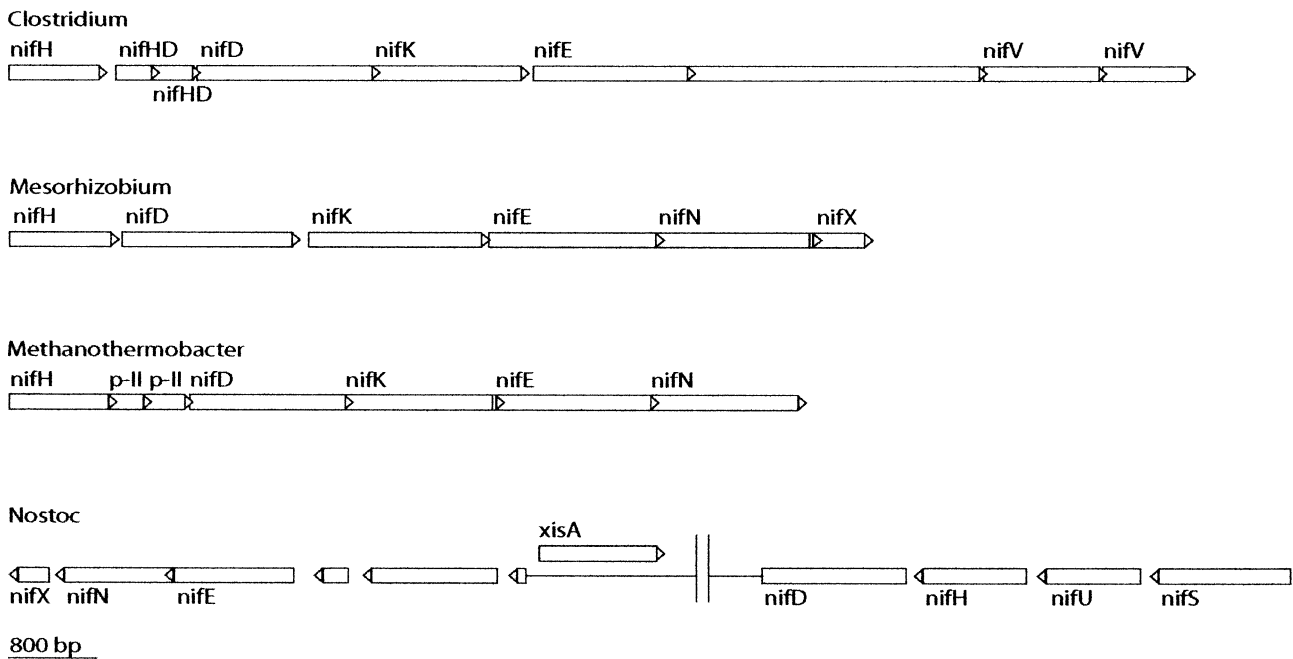
Fig. 4. Genome arrangement of *nif* genes in representative organisms from the microbial genome database (NCBI) and drawn with Artemis software (http://www.sanger.ac.uk/Software/Artemis/). Gene arrangements are shown from *Methanobacterium thermoautotrophicum* str. Delta H (NC000916), *Mesorhizobium loti* (NC 002678), *Clostridium acetobutylicum* (NC 003030), *Nostoc* sp. *Strain* 7120 (NC 003272) and *Chlorobium tepidum* TLS (NC 002932).

cations for the structure and function of planktonic microbial communities. Clearly presence of a gene in a habitat, be it nitrogenase or other 'functional' genes, must be interpreted cautiously, and may not be directly linked to the process catalysed by the expressed protein. Nonetheless, the presence of functional genes can provide markers for the movement and composition of microbial communities.

Analysis of gene expression has indicated that not all *nif* genes that are present are expressed in the environment (Noda *et al.*, 1999; Zani *et al.*, 2000), at least not at the same time. Although the regulatory cascade controlling nitrogenase gene expression is relatively well known in some enteric bacteria, little is known of the regulatory mechanisms in most prokaryotes. Interactions between microorganisms, such as the supply of photosynthate from autotrophs to heterotrophic bacteria as well as endogenous rhythms can drive the patterns of *nif* gene expression in the environment. Although it is difficult to work with mRNA from many natural sample matrices, much more work is needed in the area of gene regulation in the environment.

Genomics

The use of nitrogenase gene sequences to describe diazotroph diversity is somewhat limited by the use of *nifH*, as it is very highly conserved and the use of one gene for phylogenetic analysis can be misleading. A large number

of microbial genomes have now been sequenced, although there are still relatively few different major branches of the prokaryotes represented by genomes that contain the structural *nif* genes (*nifHDK*) in the microbial genome database. Nonetheless, it is clear that the genomic context provides information for studying phylogenetic relationships based on gene organization rather than molecular sequence of individual genes, as well as studying lateral gene transfer (Fig. 4). Furthermore, the different nitrogenase homologues can be differentiated on the basis *nif* gene arrangements (Dominic *et al.*, 2000) and by the presence and absence of specific *nif* and related genes in the genome (Fig. 4). In methanogenic Archaea, the *nifH* and *nifDK* genes are separated by *glnB* homologues, the gene for the regulatory protein PII (Fig. 4). These differences provide targets for directed molecular assays. It will be interesting to see if the gene arrangements from cultivated microorganisms are representative of the arrangements found in natural assemblages. It is likely that many surprises are in store when the genomic context of *nif* genes is investigated in the environment.

Conclusions

There are distinct distribution patterns of diazotroph phylotypes across ecosystems. Different clusters of diazotrophs are found in different types of habitats with some

similarities between physically similar habitats such as sediments and marsh soils. On the other hand, some habitats have unique and divergent populations. In some cases these diazotroph phylotypes are found even in the absence of strong N limitation.

The expanding genomic database has provided information on genome context as well as sequence information that can be used in the next generation of environmental nitrogenase studies. The use of *nifDK* gene sequences can provide additional supporting or contrasting information for evaluating *nifH* phylotypes. The gene organization data provides context that can resolve phylogenies of some of the phylotypes that are difficult to conclusively identify. However, representation of N_2 fixing microorganisms in the database of sequenced genomes is still poor.

Ultimately, expression analyses and possibly proteomics approaches will be critical for identifying the factors controlling N_2 fixation in the environment. Expression data is the key to identifying the microorganisms that are active in N_2 fixation, but more generally to understand the factors involved in controlling the distribution of nitrogenase gene-containing microorganisms in the environment. This information will help us to understand the links between diversity of functional genes and biogeochemical activity.

References

Affourtit, J., Zehr, J.P., and Paerl, H.W. (2001) Distribution of nitrogen-fixing microorganisms along the Neuse River Estuary, North Carolina. *Microb Ecol* **41:** 114–123.

Arp, D.J. (2000) The nitrogen cycle. In *Prokaryotic Nitrogen Fixation*. Triplett, E.W. (ed). Wymondham, Great Britain: Horizon Scientific Press, pp. 1–14.

Atlas, R.M., and Bartha, R. (1998) *Microbial Ecology: Fundamentals and Applications*. Menlo Park, California: Addison-Wesley Longman.

Auman, A.J., Speake, C.C., and Lidstrom, M.E. (2001) *nifH* sequences and nitrogen fixation in type I and type II methanotrophs. *Appl Environ Microbiol* **67:** 4009–4016.

Bagwell, C.E., and Lovell, C.R. (2000) Persistence of selected *Spartina alterniflora* rhizoplane diazotrophs exposed to natural and manipulated environmental variability. *Appl Environ Microbiol* **66:** 4625–4633.

Bagwell, C.E., Piceno, Y.M., Ashburne-Lucas, A., and Lovell, C.R. (1998) Physiological diversity of the rhizosphere diazotroph assemblages of selected salt marsh grasses. *Appl Environ Microbiol* **64:** 4276–4282.

Barbour, W.M., Wang, S.-P., and Stacey, G. (1992) Molecular genetics of *Bradyrhizobium* symbioses. In *Biological Nitrogen Fixation*. Stacey, G., Evans, H.J., and Burris, R.H. (eds). New York: Chapman & Hall, pp. 43–86.

Beeson, K.E., Erdner, D.L., Bagwell, C.E., Lovell, C.R., and Sobecky, P.A. (2002) Differentiation of plasmids in marine diazotroph assemblages determined by randomly amplified polymorphic DNA analysis. *Microbiology-Sgm* **148:** 179–189.

Bergholz, P.W., Bagwell, C.E., and Lovell, C.R. (2001) Physiological diversity of rhizoplane diazotrophs of the salt-meadow cordgrass, *Spartina patens*: Implications for host specific ecotypes. *Microbial Ecol* **42:** 466–473.

Bishop, P.E., Premakumar, R., Dean, D.R., Jacobsen, M.R., Chisnell, J.R., Rizzo, T.M., and Kopczynski, J. (1985) Nitrogen fixation by *Azotobacter vinelandii* strains having deletions in structural genes for nitrogenase. *Science:* 92–94.

Braun, S., Proctor, L., Zani, S., Mellon, M.T., and Zehr, J.P. (1999) Molecular evidence for zooplankton-associated nitrogen-fixing anaerobes based on amplification of the *nifH* gene. *FEMS Microbiol Ecol* **28:** 273–279.

Bull, A.T., Ward, A.C., and Goodfellow, M. (2000) Search and discovery strategies for biotechnology: The paradigm shift. *Microbiol Mol Biol Rev* **64:** 573–606,CP573.

Burgess, B.K., and Lowe, D.J. (1996) Mechanism of molybdenum nitrogenase. *Chem Rev* **96:** 2983–3011.

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

Cavalier-Smith, T. (2002) The neomuran origin of archaebacteria, the negibacterial root of the universal tree and bacterial megaclassification. *Int J Syst Evol Microbiol* **52:** 7–76.

Chen, Y.-B., Dominic, B., Mellon, M.T., and Zehr, J.P. (1998) Circadian rhythm of nitrogenase gene expression in the diazotrophic filamentous nonheterocystous cyanobacterium *Trichodesmium* sp. Strain IMS 101. *J Bacteriol* **180:** 3598–3605.

Chien, Y.-T., and Zinder, S.H. (1996) Cloning, functional organization, transcript studies, and phylogenetic analysis of the complete nitrogenase structural genes (*nifHDK2*) and associated genes in the archaeon *Methanosarcina barkeri* 227. *J Bacteriol* **178:** 143–148.

Chien, Y.T., Auerbuch, V., Brabban, A.D., and Zinder, S.H. (2000) Analysis of genes encoding an alternative nitrogenase in the archaeon *Methanosarcina barkeri* 227. *J Bacteriol* **182:** 3247–3253.

Cho, J.C., and Tiedje, J.M. (2000) Biogeography and degree of endemicity of fluorescent *Pseudomonas* strains in soil. *Appl Environ Microbiol* **66:** 5448–5456.

Des Marais, D.J., and Walter, M.R. (1999) Astrobiology: Exploring the origins, evolution, and distribution of life in the Universe. *Ann Rev Ecol Syst* **30:** 397–420.

Dicker, H., and Smith, D. (1980) Acetylene reduction (nitrogen fixation) in a Delaware, USA salt marsh. *Mar Biol* **57:** 241–250.

Dominic, B., Zani, S., Chen, Y.-B., Mellon, M.T., and Zehr, J.P. (2000) Organization of the *nif* genes of the nonheterocystous cyanobacterium *Trichodesmium* sp. IMS101. *J Phycol* **36:** 693–701.

Doolittle, W.F. (1999) Lateral genomics. *Trends Biochem Sci* **24:** M5–M8.

Dyble, J., Paerl, H.W., and Neilan, B.A. (2002) Genetic characterization of *Cylindrospermopsis raciborskii* (Cyanobacteria) isolates from diverse geographic origins based on *nifH* and *cpcBA*-IGS nucleotide sequence analysis. *Appl Environ Microbiol* **68:** 2567–2571.

Eady, R.R. (1996) Structure-function relationships of alternative nitrogenases. *Chem Rev* **96:** 3013–3030.

Egener, T., Hurek, T., and Reinhold-Hurek, B. (1998) Use of green fluorescent protein to detect expression of nif genes of *Azoarcus* sp. BH72, a grass-associated diazotroph, on rice roots. *Mol Plant–Microbe Interacts* **11:** 71–75.

Egener, T., Hurek, T., and Reinhold-Hurek, B. (1999) Endophytic expression of *nif* genes of *Azoarcus* sp strain BH72 in rice roots. *Mol Plant–Microbe Interacts* **12:** 813–819.

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

Hirsch, A.M., McKhann, H.I., Reddy, A., Liao, J., Fang, Y., and Marshall, C.R. (1995) Assessing horizontal transfer of *nifHDK* genes in eubacteria: Nucleotide sequence of *nifK* from *Frankia* Strain HFPCcI3. *Mol Biol Evol* **12:** 16–27.

Hoover, T. (2000) Control of nitrogen fixation genes in *Klebsiella pneumoniae*. In *Prokaryotic Nitrogen Fixation*. Triplett, E.W. (ed). Norfolk, England: Horizon Scientific Press, pp. 131–147.

Howard, J.B., and Rees, D.C. (1996) Structural basis of biological nitrogen fixation. *Chem Rev* **96:** 2965–2982.

Howarth, R.W., and Marino, R. (1988) Nitrogen fixation in freshwater, estuarine, and marine ecosystems. 2. Biogeochemical controls. *Limnol Oceanogr* **33:** 688–701.

Humayoun, S.B., Bano, N., and Hollibaugh, J.T. (2003) Depth distribution of microbial diversity in a meromictic lake; Mono Lake, California. *Appl Environ Microbiol* **69:** 1030–1042.

Hurek, T., Egener, T., and Reinhold-Hurek, B. (1997) Divergence in nitrogenases of *Azoarcus* spp., Proteobacteria of the beta subclass. *J Bacteriol* **179:** 4172–4178.

Hurek, T., Handley, L.L., Reinhold-Hurek, B., and Piche, Y. (2002) Azoarcus grass endophytes contribute fixed nitrogen to the plant in an unculturable state. *Mol Plant–Microbe Interacts* **15:** 233–242.

Jellison, R., and Melack, J.M. (2001) Nitrogen limitation and particulate elemental ratios of seston in hypersaline Mono Lake, California, USA. *Hydrobiologia* **466:** 1–12.

Karl, D.M., and Michaels, A.F. (1996) The Hawaiian Ocean Time-Series (Hot) and Bermuda Atlantic Time-Series Study (Bats) – Preface. *Deep-Sea Res Part Ii-Top St Oce* **43:** 127–128.

Kent, H.M., Buck, M., and Evans, D.J. (1989) Cloning and sequencing of the *nifH* gene of *Desulfovibrio gigas*. *FEMS Microbiol Lett* **61:** 73–78.

Kim, K., Zhang, Y.P., and Roberts, G.P. (1999) Correlation of activity regulation and substrate recognition of the ADP-ribosyltransferase that regulates nitrogenase activity in *Rhodospirillum rubrum*. *J Bacteriol* **181:** 1698–1702.

Klein, M., Friedrich, M., Roger, A.J., Hugenholtz, P., Fishbain, S., Abicht, H., *et al.* (2001) Multiple lateral transfers of dissimilatory sulfite reductase genes between major lineages of sulfate-reducing prokaryotes. *J Bacteriol* **183:** 6028–6035.

Lilburn, T.C., Kim, K.S., Ostrom, N.E., Byzek, K.R., Leadbetter, J.R., and Breznak, J.A. (2001) Nitrogen fixation by symbiotic and free-living spirochetes. *Science* **292:** 2495–2498.

Loveless, T.M., and Bishop, P.E. (1999) Identification of genes unique to Mo-independent nitrogenase systems in diverse diazotrophs. *Can J Microbiol* **45:** 312–317.

Lovell, C.R., Piceno, Y.M., Quattro, J.M., and Bagwell, C.E. (2000) Molecular analysis of diazotroph diversity in the rhizosphere of the smooth cordgrass, *Spartina alterniflora*. *Appl Environ Microbiol* **66:** 3814–3822.

Lovell, C.R., Friez, M.J., Longshore, J.W., and Bagwell, C.E. (2001) Recovery and phylogenetic analysis of *nifH* sequences from diazotrophic bacteria associated with dead aboveground biomass of *Spartina alterniflora*. *Appl Environ Microbiol* **67:** 5308–5314.

Madigan, M.T., Martinko, J.M., and Parker, J. (2000) *Brock Biology of Microorganisms*. Upper Saddle River, New Jersey: Prentice Hall.

Maidak, B.L., Cole, J.R., Lilburn, T.G., Parker, C.T.J.P.R.S., Farris, R.J. *et al.* (2001) The RDP-II (Ribosomal Database Project). *Nucleic Acids Res* **29:** 173–174.

Methe, B.A., Hiorns, W.D., and Zehr, J.P. (1998) Contrasts between marine and freshwater bacterial community composition: analyses of communities in Lake George, New York and six other Adirondack lakes. *Limnol Oceanogr* **43:** 368–374.

Mullins, T.D., Britschgi, T.B., Krest, R.L., and Giovannoni, S.J. (1995) Genetic comparisons reveal the same unknown bacterial lineages in Atlantic and Pacific bacterioplankton communities. *Limnol Oceanogr* **40:** 148–158.

Nesbo, C.L., L'Haridon, S., Stetter, K.O., and Doolittle, W.F. (2001) Phylogenetic analyses of two 'Archaeal' genes in *Thermotoga maritima* reveal multiple transfers between Archaea and Bacteria. *Mol Biol Evol* **18:** 362–375.

Noda, S., Ohkuma, M., Usami, R., Horikoshi, K., and Kudo, T. (1999) Culture-independent characterization of a gene responsible for nitrogen fixation in the symbiotic microbial community in the gut of the termite *Neotermes koshunensis*. *Appl Environ Microbiol* **65:** 4935–4942.

Ohkuma, M., Noda, S., Usami, R., Horikoshi, K., and Kudo, T. (1996) Diversity of nitrogen fixation genes in the symbiotic intestinal microflora of the termite *Reticulitermes speratus*. *Appl Environ Microbiol* **62:** 2747–2752.

Ohkuma, M., Noda, S., and Kudo, T. (1999) Phylogenetic diversity of nitrogen fixation genes in the symbiotic microbial community in the gut of diverse termites. *Appl Environ Microbiol* **65:** 4926–4934.

Olson, J.B., Litaker, R.W., and Paerl, H.W. (1999) Ubiquity of heterotrophic diazotrophs in marine microbial mats. *Aquat Micr Ecol* **19:** 29–36.

Pace, N.R. (1997) A molecular view of microbial diversity and the biosphere. *Science* **276:** 734–740.

Paerl, H.W. (1988) Nuisance phytoplankton blooms in coastal, estuarine, and inland waters. *Limnol Oceanogr* **33:** 823–847.

Piceno, Y.M., and Lovell, C.R. (2000a) Stability in natural bacterial communities. II. Plant resource allocation effects on rhizosphere diazotroph assemblage composition. *Microb Ecol* **39:** 41–48.

Piceno, Y.M., and Lovell, C.R. (2000b) Stability in natural bacterial communities. I. Nutrient addition effects on rhizosphere diazotroph assemblage composition. *Microb Ecol* **39:** 32–40.

Poly, F., Monrozier, L.J., and Bally, R. (2001a) Improvement in the RFLP procedure for studying the diversity of *nifH*

genes in communities of nitrogen fixers in soil. *Res Microbiol* **152:** 95–103.

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

Postgate, J.R., and Eady, R.R. (1988) The evolution of biological nitrogen fixation. In *Nitrogen Fixation: Hundred Years After*. Bothe, H., de Bruijn, F.J, and Newton, W.E. (eds). Stuttgart: Gustav Fischer, pp. 31–40.

Rappe, M.S., Connon, S.A., Vergin, K.L., and Giovannoni, S.J. (2002) Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. *Nature* **418:** 630–633.

Redfield, A.C. (1958) The biological control of chemical factors in the environment. *Am Scientist* **46:** 205–222.

Ribbe, M., Gadkari, D., and Meyer, O. (1997) N_2 fixation by *Streptomyces thermoautotrophicus* involves a molybdenum-dinitrogenase and a manganese-superoxide oxidoreductase that couple N_2 reduction to the oxidation of superoxide produced from O_2 by a molybdenum-CO dehydrogenase. *J Biol Chem* **272:** 26627–26633.

Rosado, A.S., Duarte, G.F., Seldin, L., and VanElsas, J.D. (1998) Genetic diversity of *nifH* gene sequences in *Paenibacillus azotofixans* strains and soil samples analyzed by denaturing gradient gel electrophoresis of PCR-amplified gene fragments. *Appl Environ Microbiol* **64:** 2770–2779.

Shaffer, B.T., Widmer, F., Porteous, L.A., and Seidler, R.J. (2000) Temporal and spatial distribution of the *nifH* gene of N_2 fixing bacteria in forests and clearcuts in western Oregon. *Microb Ecol* **39:** 12–21.

Simonet, P., Grosjean, M.C., Misra, A.K., Nazaret, S., Cournoyer, B., and Normand, P. (1991) *Frankia* genus-specific characterization by polymerase chain reaction. *Appl Environ Microbiol* **57:** 3278–3286.

Steppe, T.F., and Paerl, H.W. (2002) Potential N_2 fixation by sulfate-reducing bacteria in a marine intertidal microbial mat. *Aquat Micr Ecol* **28:** 1–12.

Steppe, T.F., Olson, J.B., Paerl, H.W., Litaker, R.W., and Belnap, J. (1996) Consortial N_2 fixation – A strategy for meeting nitrogen requirements of marine and terrestrial cyanobacterial mats. *FEMS Microbiol Ecol* **21:** 149–156.

Steppe, T.F., Pinckney, J.L., Dyble, J., and Paerl, H.W. (2001) Diazotrophy in modern marine Bahamian stromatolites. *Microb Ecol* **41:** 36–44.

Strunk, O. and Ludwig, W. (1993) The ARB-Project: A software environment for sequence data (Munich, Germany). *http://www.arb-home.de*.

Tamas, I., Svircev, Z., and Andersson, S.G.E. (2000) Determinative value of a portion of the *nifH* sequence for the genera *Nostoc* and *Anabaena* (Cyanobacteria). *Current Microbiol* **41:** 197–200.

Tan, Z.Y., Hurek, T., Vinuesa, P., Muller, P., Ladha, J.K., and Reinhold-Hurek, B. (2001) Specific detection of *Bradyrhizobium* and *Rhizobium* strains colonizing rice (*Oryza sativa*) roots by 16S–23S ribosomal DNA intergenic spacer-targeted PCR. *Appl Environ Microbiol* **67:** 3655–3664.

Thiel, T. (1993) Characterization of genes for an alternative nitrogenase in the cyanobacterium *Anabaena variabilis*. *J Bacteriol* **175:** 6276–6286.

Thiel, T., Lyons, E.M., Erker, J.C., and Ernst, A. (1995) A second nitrogenase in vegetative cells of a heterocyst-forming cyanobacterium. *Proc Natl Acad Sci USA* **92:** 9358–9362.

Ueda, T., Suga, Y., Yahiro, N., and Matsuguchi, T. (1995a) Remarkable N_2 fixing bacterial diversity detected in rice roots by molecular evolutionary analysis of *nifH* gene sequences. *J Bacteriol* **177:** 1414–1417.

Ueda, T., Suga, Y., Yahiro, N., and Matsuguchi, T. (1995b) Genetic diversity of N_2 fixing bacteria associated with rice roots by molecular evolutionary analysis of a *nifD* library. *Can J Microbiol* **41:** 235–240.

Vitousek, P.M. (1999) Nutrient limitation to nitrogen fixation in young volcanic sites. *Ecosystems* **2:** 505–510.

Vitousek, P.M., and Howarth, R.W. (1991) Nitrogen limitation on land and in the sea: How can it occur? *Biogeochemistry* **13:** 87–115.

Wang, S., Chen, J., and Johnson, J.L. (1988) The presence of five *nifH*-like sequences in *Clostridium pasteurianum*: sequence divergence and transcription properties. *Nucleic Acids Res* **16:** 439–454.

Wu, J., Sunda, W., Boyle, E.A., and Karl, D.M. (2000) Phosphate depletion in the western North Atlantic Ocean. *Science* **289:** 759–762.

Young, J.P.W. (1992) Phylogenetic classification of nitrogen-fixing organisms. In *Biological Nitrogen Fixation*. Stacey, G., Evans, H.J., and Burris, R.H. (eds). New York: Chapman & Hall, pp. 43–86.

Zani, S., Mellon, M.T., Collier, J.L., and Zehr, J.P. (2000) Expression of *nifH* genes in natural microbial assemblages in Lake George, New York detected with RT-PCR. *Appl Environ Microbiol* **66:** 3119–3124.

Zehr, J.P., and Capone, D.G. (1996) Problems and promises of assaying the genetic potential for nitrogen fixation in the marine environment. *Microb Ecol* **32:** 263–281.

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

Zehr, J.P., Mellon, M., Braun, S., Litaker, W., Steppe, T., and Paerl, H.W. (1995) Diversity of heterotrophic nitrogen fixation genes in a marine cyanobacterial mat. *Appl Environ Microbiol* **61:** 2527–2532.

Zehr, J.P., Mellon, M.T., and Hiorns, W.D. (1997) Phylogeny of cyanobacterial *nifH* genes: evolutionary implications and potential applications to natural assemblages. *Microbiology* **143:** 1443–1450.

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

Zehr, J.P., Waterbury, J.B., Turner, P.J., Montoya, J.P., Omoregie, E., Steward, G.F., *et al.* (2001) Unicellular cyanobacteria fix N_2 in the subtropical North Pacific Ocean. *Nature* **412:** 635–638.