

The Evolutionary History of Nitrogen Fixation, as Assessed by *NifD*

Brian J. Henson, Linda E. Watson, Susan R. Barnum

Department of Botany, Miami University, Oxford, OH 45056, USA

Received: 12 May 2003 / Accepted: 23 October 2003

Abstract. The evolutionary history of nitrogen fixation has been vigorously debated for almost two decades. Previous phylogenetic analyses of nitrogen fixation genes (*nif*) have shown support for either evolution by vertical descent or lateral transfer, depending on the specific *nif* gene examined and the method of analyses used. The debate centers on the placement and monophyly of the cyanobacteria, proteobacteria, and Gram-positive bacteria (actinobacteria and firmicutes). Some analyses place the cyanobacteria and actinobacteria within the proteobacteria, which suggests that the *nif* genes have been laterally transferred since this topology is incongruent with ribosomal phylogenies, the standard marker for comparison. Other *nif* analyses resolve and support the monophyly of the cyanobacteria, proteobacteria, and actinobacteria, supporting vertical descent. We have revisited these conflicting scenarios by analyzing *nifD* from an increased number of cyanobacteria, proteobacteria, and Gram-positive bacteria. Parsimony analyses of amino acid sequences and maximum likelihood analysis of nucleic acid sequences support the monophyly of the cyanobacteria and actinobacteria but not the proteobacteria, lending support for vertical descent. However, distance analysis of nucleic acid sequences placed the actinobacteria within the proteobacteria, supporting lateral transfer. We discuss evidence for both vertical descent and lateral transfer of nitrogen fixation.

Key words: *nifD* — Nitrogen fixation — Lateral transfer — Vertical descent — *nif* phylogeny

Introduction

Nitrogen fixation (or diazotrophy) is the process of converting atmospheric nitrogen (N_2) to reduced forms such as ammonia (NH_3) (Postgate 1982). A small amount of atmospheric dinitrogen is reduced by lightning; however, the majority is reduced by prokaryotes (Postgate 1982; Sprent and Sprent 1990). The enzyme complex nitrogenase is responsible for nitrogen fixation. The subunits of nitrogenase are encoded by the *nifHDK* operon. Nitrogenase is composed of two components, dinitrogenase reductase (iron protein) and dinitrogenase (molybdenum-iron protein). Dinitrogenase reductase is composed of two identical subunits that are encoded by *nifH* (Mevarech et al. 1980), while dinitrogenase is a tetramer composed of two subunits encoded by *nifD* (Lammers and Haselkorn 1984) and two subunits encoded by *nifK* (Mazur and Chui 1982). Dinitrogenase reductase contains an iron-sulfur (4F-4S) cofactor, which binds the subunits of nitrogenase, and is responsible for mediating the ATP-dependent transfer of electrons to the dinitrogenase tetramer. Dinitrogenase binds atmospheric dinitrogen (N_2) and is responsible for the transfer of electrons to it (Postgate 1982).

Other nitrogen fixation operons, in addition to *nifHDK*, have been identified in the heterocystous cyanobacterium, *Nostoc* sp. Strain PCC 7120 (Mazur and Chui 1982; Rice et al. 1982; Lammers and Has-

elkorn 1984). These include *nif*BSU, *fdxN* (Mulligan et al. 1988; Mulligan and Haselkorn 1989), and *nif*ENXW (Borthakur et al. 1990; Haselkorn and Bukema 1992). *NifB*, *nifN*, and *nifE* are involved in molybdenum-iron cofactor synthesis. *FdxN* encodes a bacterial-type ferredoxin of unknown function, and the functions of *nifS*, *nifU*, *nifX*, and *nifW* remain unknown. Recent studies have suggested that some *nif* genes may have arisen via paralogous gene duplication (Fani et al. 2000). In addition to the *nif* gene family listed above, several alternative nitrogenases have been found and these include Mo-dependent *nif2*, Va-dependent *vnf*, and Fe-dependent *anf* operons (Bishop and Premakumar 1992).

The ability to fix nitrogen is widely distributed in distantly related members of the Eubacteria and the Archaea (Young 1992); however, it is not ubiquitous in occurrence. The evolutionary history of the nitrogen fixation has been debated for some time. The debate has focused on how to explain the random distribution of nitrogen fixation in distantly related lineages of prokaryotes. Several alternative hypotheses have been proposed to explain this haphazard distribution. The first hypothesis is that nitrogen fixation arose once in evolutionary history but has been transferred laterally (i.e., horizontally) to various lineages (Normand and Bousquet 1989). A second hypothesis predicts that it arose early in the evolutionary history of prokaryotes and was at one time ubiquitous among them, but has since been lost by many lineages and retained by a few distantly related ones (i.e., vertical descent, accompanied by multiple losses) (Young 1992; Normand et al. 1992; Hirsch et al. 1995). A third view is that it arose multiple times through convergent evolution (Postgate and Eady 1988).

To evaluate the evolution of nitrogen fixation, we and others (Normand and Bousquet 1989; Normand et al. 1992; Hirsch et al. 1995) have examined the phylogenies of *nif* genes. We utilized published ribosomal RNA phylogenies (Woese 1987; Olsen and Woese 1993; Olsen et al. 1994) as an independent framework for comparison because rRNA is universally present, highly conserved, believed to be representative of the organismal phylogeny, and not believed to be laterally transferred (Woese 1987; Olsen and Woese 1993; Olsen et al. 1994). Thus, rRNA phylogenies in all likelihood, represent true organismal phylogenies and, therefore, are the most appropriate comparison for exploring alternative hypotheses involving lateral gene transfer. If the *nif* and the rRNA phylogenies have generally congruent topologies, then vertical descent may be supported. If, on the other hand, the *nif* and rRNA phylogenies have incongruent topologies, then lateral transfer of *nif* may have occurred. Convergence, or multiple origins of *nif* genes, could be indicated by incongruence, although we would also expect significant

divergence in nucleotide and amino acid sequences if this were the case. This is in sharp contrast to the high sequence similarity and the conserved organization of *nif* genes found in nitrogen fixing organisms. Another possibility to explain incongruence is the inadvertent analysis of paralogous gene copies that arose through duplication. Therefore, our study focuses on distinguishing horizontal transfer from vertical descent of *nif* genes within the Eubacteria. In this paper, we reexamine the evolutionary history of nitrogen fixation by analyzing *nifD* from an increased number of cyanobacteria, proteobacteria, and Gram-positive bacteria. We analyzed only molybdenum containing *nifD* genes, and excluded all alternative nitrogenase genes, i.e., *anfD* and *vnfD*. Our objective was to compare the topology of our *nifD* phylogeny to 16S rRNA phylogenies to assess the evolutionary history of *nifD* and nitrogen fixation within the Eubacteria.

Materials and Methods

Fifty-seven *nifD* nucleic acid and inferred amino acid sequences were used in this study (Table 1). Inferred amino acid sequences were initially aligned using Clustal W (Thompson et al. 1994) with gaps inserted for optimal alignment. The amino acid alignment was then adjusted manually using MacClade 4.0 (Maddison and Maddison 2000). The nucleotide alignment was made using Codon Align 1.0 (Barry Hall, University of Rochester), which constructs the alignment based on the amino acid alignment, with gaps inserted between codons and not within them. The amino acid alignment was then adjusted manually using MacClade 4.0 (Maddison and Maddison 2000).

The amino acid data matrix was analyzed using parsimony with PAUP* 4.0b10 (Swofford 2002). All trees were rooted with outgroup analysis using six archaean representatives (Table 1). Analyses were performed with the user-defined stepmatrix "PROTPARS," which is equivalent to Felsenstein's PROTPARS from PHYLIP (Felsenstein 1981). Parsimony analysis was conducted using the heuristic search option, with gaps treated as missing data, trees obtained using stepwise addition, and the tree-bisection-reconnection (TBR) branch swapping with random sequence addition for 100 replicates, steepest descent not in effect, maxtrees set at 50,000, branches collapsed if the maximum length was zero, multrees option in effect, and no topological constraints enforced. Three parsimony analyses were performed: (1) all characters included, (2) constant and uninformative characters excluded, and (3) constant and uninformative characters plus 51 characters that correspond to an insertion found in *nifD* genes of *Clostridium*, *Methanosarcina*, and *Chlorobium tepidum*. Bootstrap values were calculated for 500 replicates to evaluate branch support using the parsimony options listed above (Bremer 1994; Felsenstein 1985; Huelsenbeck et al. 1995).

Analysis of the nucleic acid data matrix was performed using maximum likelihood (ML) criteria with PAUP* 4.0b10 (Swofford 2002). We determined the evolutionary model that best described our data using Modeltest 3.06 (Posada and Crandall 1998), which was the F81 model (Felsenstein 1981) with a likelihood score of 22106.18947. Trees were rooted with outgroup analysis using *nifD* sequences from six archaean representatives (Table 1). ML analysis was conducted using the heuristic search option, with the likelihood settings corresponding to the F81 model (Felsenstein 1981). ML settings were as follows: no molecular clock enforced, starting branch lengths determined via the Rogers-Swofford approximation method, trees with likelihoods that were 5% or further from the

Table 1. Eubacterial and archaeal *nifD* sequences used in this study with GenBank accession numbers listed

Organism	Accession no.
Cyanobacteria	
<i>Anabaena</i> sp. PCC 7108	AF442505
<i>Anabaena cylindrica</i> PCC 7122	AF442506
<i>Anabaena</i> sp. ATCC 33047	AF442507
<i>Calothrix desertica</i> PCC 7102	AF442501
<i>Chlorogloeopsis fritschii</i> PCC 6718	AF442510
<i>Cylindrospermum stagnale</i> PCC 7417	AF442508
<i>Fischerella musicola</i> PCC 7414	AF442514
<i>Fischerella</i> sp. UTEX 1931	U49514
<i>Nodularia spumigena</i> PCC 73104	AF442509
<i>Nostoc commune</i> UTEX 584	L23514
<i>Nostoc</i> sp. PCC 7423	AF442503
<i>Nostoc</i> sp. PCC 6720	AF442502
<i>Nostoc</i> sp. PCC 7120	AF442504
<i>Scytonema hofmanni</i> PCC 7110	AF442511
<i>Cyanothece</i> sp. PCC 8801	U22146
<i>Cyanothece</i> sp. ATCC 51142	AF003337
<i>Plectonema boryanum</i>	D00666
<i>Trichodesmium</i> IMS101	AF016484
Proteobacteria	
<i>Azospirillum brasilense</i> [α]	M64344
<i>Bradyrhizobium</i> sp. Tv2a-2 [α]	AF285776
<i>Bradyrhizobium</i> sp. iwc91-2 [α]	AF285777
<i>Bradyrhizobium elkanii</i> [α]	AF484268
<i>Bradyrhizobium japonicum</i> [α]	AF484262
<i>Bradyrhizobium liaoningense</i> [α]	AF484263
<i>Bradyrhizobium</i> CJ3.3 [α]	AF484276
<i>Gluconacetobacter diazotrophicus</i> [α]	AF030414
<i>Mesorhizobium</i> sp. Lc3-2 [α]	AF411040
<i>Mesorhizobium</i> sp. Lo7-12 [α]	AF411039
<i>Mesorhizobium</i> sp. Lo9-4 [α]	AF411038
<i>Mesorhizobium loti</i> R7A [α]	BA000012
<i>Rhizobium etli</i> [α]	L13618
<i>Rhizobium</i> sp. ANU240 [α]	M26962
<i>Rhizobium</i> sp. ANU289 [α]	X01139
<i>Rhizobium</i> sp. IRc78 [α]	M10203
<i>Sinorhizobium meliloti</i> [α]	AL591688
<i>Alcaligenes faecalis</i> [β]	X95565
<i>Azoarcus</i> sp. BH72 [β]	AF200742
<i>Burkholderia</i> sp. [β]	AF194084
<i>Burkholderia</i> sp. STM678 [β]	AJ302315
<i>Herbaspirillum seropedicae</i> [β]	Z54207
<i>Rhodobacter capsulatus</i> [α]	M15270
<i>Acidithiobacillus ferrooxidans</i> [γ]	M15238
<i>Azotobacter vinelandii</i> [γ]	X06886
<i>Klebsiella pneumoniae</i> [γ]	Y00316
Gram-positive bacteria	
Actinobacteria (high G + C)	
<i>Frankia alni</i>	L41344
<i>Frankia</i> EulK1	U53362
<i>Frankia</i> sp. FaC1	FSU53363
<i>Frankia</i> sp. Mrp182	AY115490
Firmicutes (Low G + C)	
<i>Clostridium pasteurianum</i>	Y00155
<i>Clostridium acetobutylicum</i>	NC_003030
Green-sulfur bacteria	
<i>Chlorobium tepidum</i>	AE006470
Archaea outgroups	
<i>Methanococcus thermolithotrophicus</i>	X13830
<i>Methanococcus maripaludis</i>	U75887
<i>Methanothermobacterium thermautotrophicum</i>	X87971
<i>Methanosarcina acetivorans</i>	AE010299
<i>Methanosarcina barkeri</i>	NZ_AAAR00000000
<i>Methanosarcina mazei</i>	NC_003901

target score were rejected, branch-length optimization, equaled one-dimensional Newton Raphson with pass where the limit = 20 and $\delta = 1e-06$, starting trees obtained via stepwise addition, sequence addition as-is, one tree held at each step during stepwise addition, TBR, steepest descent in effect, maxtrees set at 50,000, branches collapsed if the branch lengths equal to or less than $1e-08$, multrees option in effect, and topological constraints not enforced. ML analysis was performed with the third codon excluded, to remove noise resulting from saturation at that position, as well as with and without the ~ 154 -bp insert excluded. Bootstrap values were calculated for 500 replicates to evaluate branch support using parsimony criteria (Bremer 1994; Felsenstein 1985; Huel- senbeck et al. 1995).

Distance analysis of the nucleic acid data matrix was performed using the neighbor-joining method (Saitou and Nei 1987), with all characters included and the DNA distance measure set to LogDet/paralinear to correct for base compositional bias (Lake 1994; Lockhart et al. 1994).

Results

NifD Amino Acid Sequences

The aligned *nifD* amino acid data matrix was 599 residues in length and consisted of 349 variable characters. Parsimony analysis with all characters included generated 38 equally most parsimonious trees that were 4389 steps long, with a consistency index (CI) of 0.54, a rescaled consistency index (RC) of 0.39, and a retention index (RI) of 0.71. The 38 most parsimonious trees (MPTs) represented or six islands of similar topology (Fig. 1). The major topological difference between the islands is the placement of *Acidithiobacillus ferrooxidans* and a clade containing *Alcaligenes faecalis*, *Azotobacter vinelandii*, *Azoarcus* sp. BH72, and *Klebsiella pneumoniae*. The first island is composed of four trees (Fig. 1A) with a branch containing *A. faecalis*, *A. vinelandii*, *Azoarcus* sp. BH72, *K. pneumoniae* sister to the cyanobacteria, and *Acidithiobacillus ferrooxidans* sister to the cyanobacteria and proteobacteria. The second island is composed of 19 trees (Fig. 1B), with the cyanobacteria and proteobacteria as sister, with *A. ferrooxidans* as sister to the larger clade. Islands 3 through 6, composed of 12, 1, 1, and 1 trees, respectively, place the cyanobacteria and proteobacteria as sister, with *A. ferrooxidans* embedded within the proteobacteria, and *A. faecalis*, *A. vinelandii*, *Azoarcus* sp. BH72, and *K. pneumoniae* as sister (Fig. 1C). The islands vary slightly at the terminal branches, but the overall topology of all MPTs is similar. A strict consensus tree of the 38 MPTs from the six islands is presented in Fig. 1D. This tree consists of seven major branches. The cyanobacteria and proteobacteria occur in a single clade, with the cyanobacteria supported as monophyletic and the proteobacteria occurring on several unresolved clades. Neither the α , β , nor γ proteobacterial subgroups are supported as monophyletic.

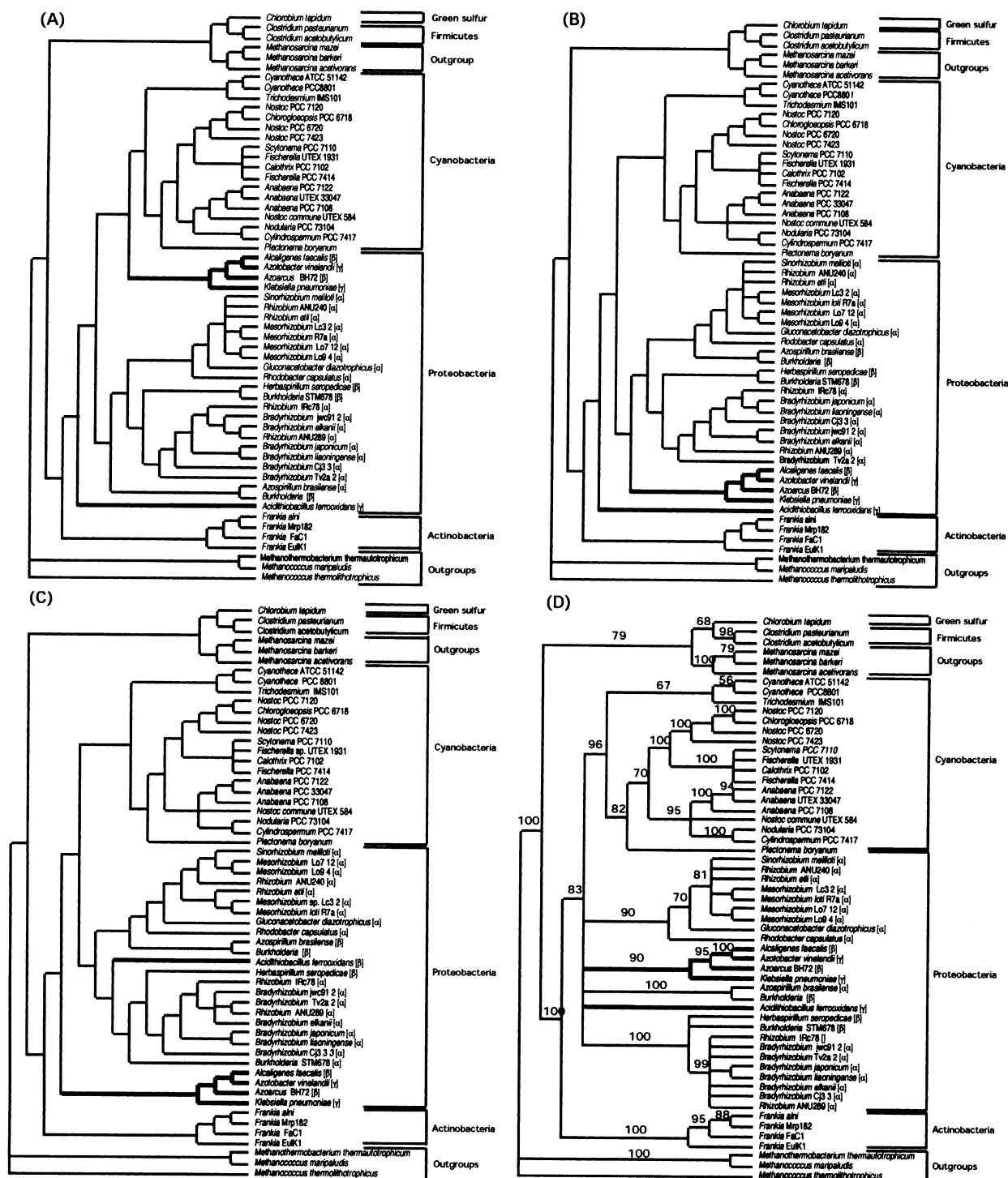


Fig. 1. Trees generated from parsimony analysis of *nifD* amino acid sequences. The taxonomic groups are indicated; in addition, the proteobacterial subgroups are in brackets. Branches in boldface are those taxa that occur on different islands (discussed in the text). **A** Strict consensus tree of the four equally parsimonious trees in

island 1. **B** Strict consensus tree of the 19 equally parsimonious trees in island 2. **C** Strict consensus tree generated from the 15 equally parsimonious trees generated from islands 3 through 6. **D** Strict consensus tree of all 38 equally parsimonious trees. Numbers above branches indicate bootstrap values.

In all MPTs of amino acid sequences (Fig. 1), the Gram-positive bacteria (firmicutes and actinobacteria) are not supported as monophyletic. The high G+C Gram-positive bacteria and *Frankia* strains (actinobacteria) are monophyletic and sister to the

proteobacteria and cyanobacteria. The low G+C Gram-positive bacteria, *Clostridia* species (firmicutes), form a clade sister to the green sulfur bacterium, *Chlorobium tepidum*. All additional parsimony analyses of the amino acid data matrix (i.e., with and

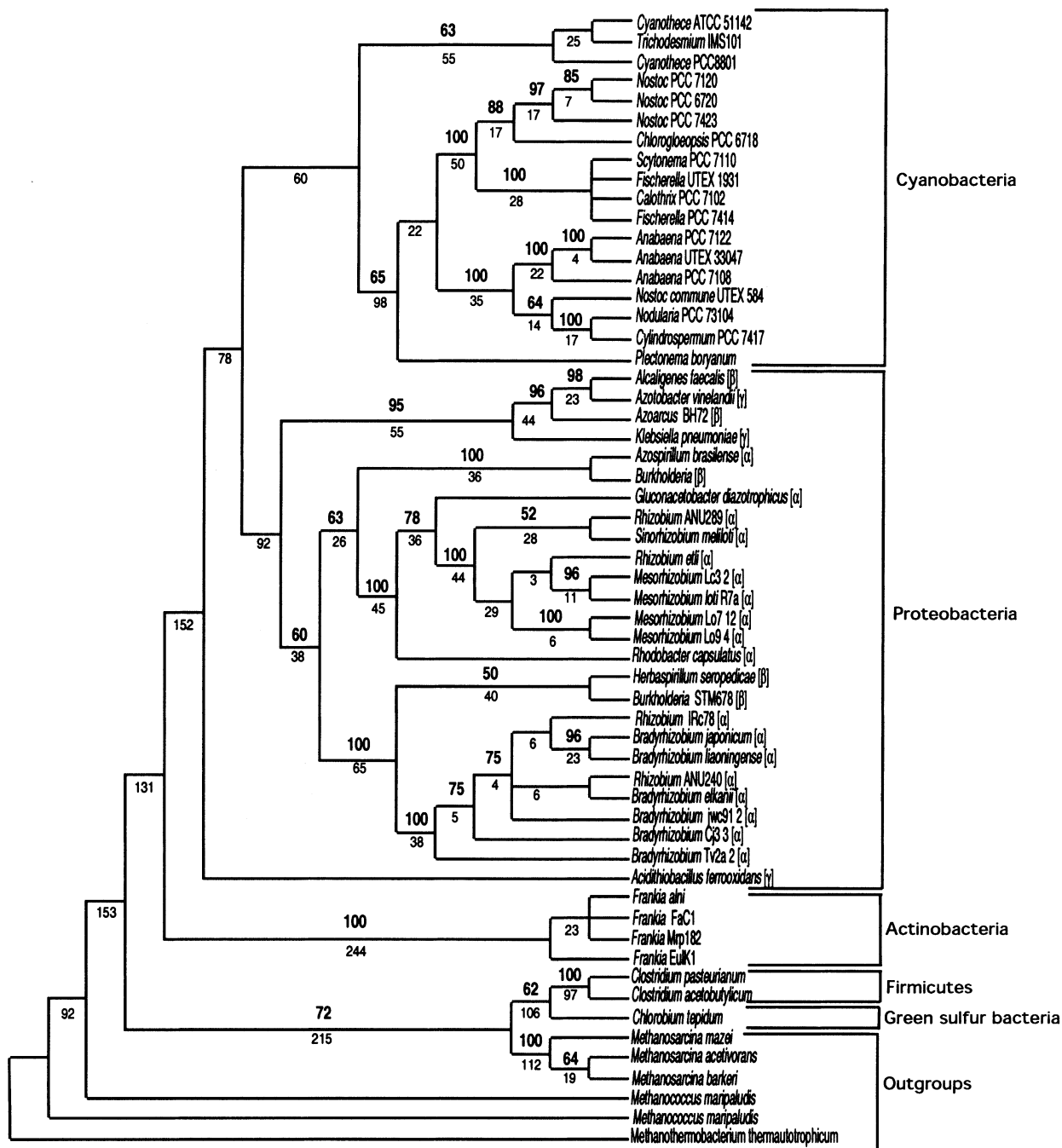


Fig. 2. Maximum likelihood tree of *nifD* nucleic acid sequences that had an $-\ln L$ score of 20239.65577. The different groups of bacterial taxa are bracketed. The proteobacterial subgroups are in brackets. Numbers in boldface above the branches are bootstrap values and numbers below the branches indicate branch lengths.

without constant and uninformative characters and the 51 amino acid residue insertion) resulted in trees that are virtually identical in topology to those trees presented in Fig. 1.

NifD Nucleic Acid Analysis

The aligned *nifD* nucleic acid data matrix was 1794 bp in length, with the 598 third codon positions excluded from the analysis. ML analysis converged on a single

tree with a $-\ln L$ score of 20239.65577 (Fig. 2) and resolved four major clades. The cyanobacteria are supported as monophyletic and are sister to the proteobacteria. However, ML analysis failed to support the monophyly of the α , β , or γ subgroups of the proteobacteria. Sister to the cyanobacteria and proteobacteria is *Acidithiobacillus ferrooxidans*. The Gram-positive bacteria (actinobacteria and firmicutes) are not resolved as monophyletic, while the actinobacteria are monophyletic and sister to the

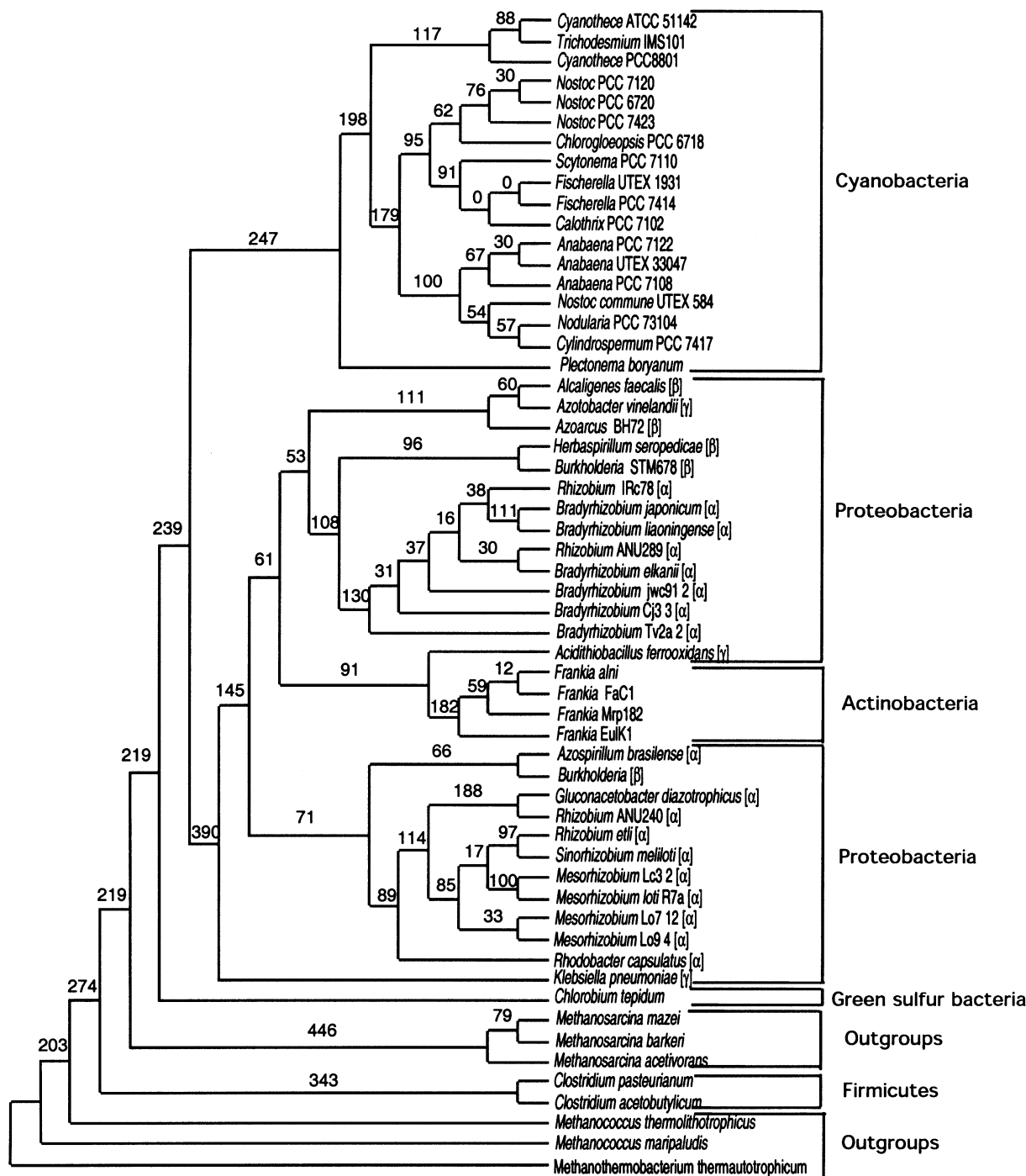


Fig. 3. Distance tree of *nifD* nucleic acid sequences created with the neighbor-joining method and LogDet/paralinear DNA distances. The different groups of bacterial taxa are bracketed. The proteobacterial subgroups are in brackets. Numbers above the branches indicate branch lengths.

cyanobacteria and proteobacteria. *Clostridium* species occur on a branch with *Chlorobium tepidum* and *Methanosarcina*.

The neighbor-joining tree (Fig. 3), constructed using LogDet/paralinear distances, resulted in a tree with a topology similar to that of the ML analysis of DNA sequences (Fig. 2) and parsimony analysis of amino acid sequences (Fig. 1); however, *Frankia* is embedded within the proteobacteria. Distance anal-

ysis also differed from the other analyses (Figs. 1 and 2) in that *Chlorobium tepidum*, *Clostridium* species, and *Methanosarcina* species are not placed together.

Discussion

The relationships among the cyanobacteria, proteobacteria, and Gram-positive bacteria (actinobacteria and firmicutes) have not been fully resolved from

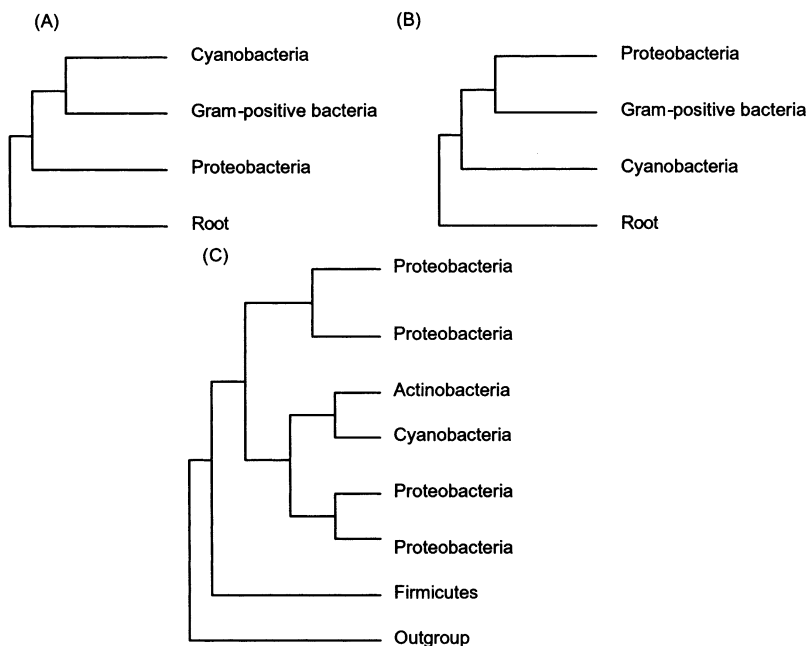


Fig. 4. Published trees redrawn. **A** 16S rRNA tree modified from Woese (1987). **B** 16S rRNA tree modified from Olsen et al. (1993). **C** Phylogeny depicting lateral transfer from the proteobacteria to the cyanobacteria and actinobacteria. Adapted from Hirsch et al. (1995).

analyses of ribosomal RNAs (Hirsch et al. 1995). Phylogenies based on 16S rRNA differ in the relationships between these eubacterial lineages, with Woese (1987) placing the cyanobacteria and Gram-positive bacteria as sister to each other with the proteobacteria at the base of the tree (Fig. 4A), whereas Olsen et al. (1994) places the Gram-positive bacteria and the proteobacteria as sister to each other with the cyanobacteria at the base of the tree (Fig. 4B). Although these topologies differ, they are consistent in the resolution of these eubacterial lineages as monophyletic and distinct. Therefore, it has been suggested that the relationship among these eubacterial lineages may best be described as an unresolved trichotomy (Hirsch et al. 1995). However, these phylogenies are still useful for comparison because they resolve the three lineages as monophyletic. Thus, incongruence between the *nif* phylogenies and the 16S rRNA phylogenies may indicate lateral gene transfer. For example, if lateral transfer of *nif* had occurred, we would not expect to resolve these major lineages as monophyletic. Rather, we would expect that members of one lineage would be placed among members of another lineage (Fig. 4C), which has been reported by others (Normand and Bousquet 1989; Normand et al. 1992; Hirsch et al. 1995). *Nif*H, -D, and -K have all been examined to some extent and have produced conflicting results supporting both lateral transfer and vertical descent.

Analyses of partial *nif*H sequences that support lateral transfer have topologies that were incongruent with a 16S rRNA-based phylogeny and placed the cyanobacteria and actinobacteria within the proteobacteria (Normand and Bousquet 1989; Normand et al. 1992; Hirsch et al. 1995; Ueda et al. 1995;

Kessler et al. 1997). This suggests that nitrogen fixation was laterally transferred from an ancestral proteobacterium to the most recent common ancestors of cyanobacteria and actinobacteria (Normand and Bousquet 1989; Normand et al. 1992; Hirsch et al. 1995). However, other *nif*H analyses resolved the cyanobacteria and proteobacteria as sister to each other with the actinobacteria at the base of the tree (Hennecke et al. 1985; Hirsch et al. 1995; Zehr et al. 1998, 2000), consistent with the topology of the 16S rRNA phylogeny supporting vertical descent of *nif*H.

Previous *nif*D phylogenies also supported lateral transfer because the cyanobacteria were placed within the proteobacteria (Normand and Bousquet 1989; Hirsch et al. 1995; Kessler et al. 1997). The *nif*D phylogeny differs from *nif*H in that the actinobacteria and cyanobacteria do not occur together, and actinobacteria are placed as sister to the proteobacteria and cyanobacteria clade (Normand et al. 1992; Kessler et al. 1997). The failure of *nif*D to resolve the evolution of nitrogen fixation led some investigators to speculate that it simply does not provide sufficient resolution (Normand et al. 1992; Hirsch et al. 1995).

*Nif*K amino acid sequences have also been analyzed with parsimony and distance methods to resolve the issue (Hirsch et al. 1995). Parsimony analysis resulted in a phylogeny that predicted vertical descent, with the actinobacteria and cyanobacteria as sister to the proteobacteria (Hirsch et al. 1995). In contrast, distance analysis resulted in a phylogeny that supported lateral transfer (Fig. 4C), with the actinobacteria and cyanobacteria sister to each other and embedded within the proteobacteria (Hirsch et al. 1995). Thus, conflicting results may be a combination of three different *nif* genes being analyzed

with different methods of phylogenetic analysis, none of which are model-based methods such as ML.

Our analyses of *nifD* (Figs. 1–3), consistently produced topologies that are congruent with 16S rRNA phylogenies, and with the cyanobacteria, proteobacteria, and actinobacteria fully resolved via reciprocal monophyly and, thus support vertical descent. Our analysis did not result in members of one lineage being placed within or among members of another lineage, consistent with 16S rRNA phylogenies supporting vertical descent.

Parsimony analysis of amino acid sequences (Fig. 1) and ML analyses of DNA sequences (Fig. 2), place the firmicutes (*Clostridium*), *Methanosarcina*, and *Chlorobium tepidum* together at the base of the tree; however, distance analysis of DNA sequences (Fig. 3) places them in two clades at the base of the tree. The placement of these taxa together is unexpected; since the firmicutes and *Chlorobium tepidum* are Eubacteria, and *Methanosarcina* is a member of the Archaea. Although the placement of these taxa together may be unexpected, analyses of *nifH* also place *Clostridium* with *Methanosarcina* (Zehr et al. 1998, 2000). While this might be attributed to long-branch attraction, *Clostridium*, *Methanosarcina*, and *Chlorobium tepidum* all possess an insertion encoding approximately 50 amino acids not found in any of the other taxa examined. However, this 50-amino acid insert is not conserved in these taxa. Distance analysis using the LogDet/paralinear distances, which compensates for base compositional bias (Lake 1994; Lockhart et al. 1994), did not place these taxa together (Fig. 3). This may suggest that the *nifD* genes of these taxa share a common evolutionary history due to lateral gene transfer. It is also possible that the placement of these two taxa together is due to their *nifD* sequences being ancient paralogous copies that arose prior to the separation of the Eubacteria and the Archaea (Young 1990, 1992, 2000; Leigh 2001).

Another discrepancy in our results is the placement of *Frankia* in the parsimony analysis of amino acid sequences (Fig. 1) and ML analysis of DNA sequences (Fig. 2), which place them as sister to the larger clade of cyanobacteria and proteobacteria, in contrast to our distance analysis of DNA sequences (Fig. 3) which places it within the proteobacteria. The placement of *Frankia* within the proteobacteria (Fig. 3) is consistent with lateral transfer, but its placement as sister to the cyanobacteria and proteobacteria (Fig. 2) is consistent with vertical descent. Although there is conflict regarding the placement of *Frankia* in the distance tree (Fig. 3), it is more likely that its placement as sister to the cyanobacteria and proteobacteria is correct, supporting vertical descent (Figs. 1 and 2), since this is congruent with 16S rRNA (Woese 1987; Olsen and Woese 1993; Olsen

et al. 1994), and the increased probability that parsimony and likelihood methods are superior phylogenetic analyses for more accurately resolving evolutionary relationships.

Our results are consistent with Zehr et al. (1998, 2000) in the placement of the proteobacteria and cyanobacteria as sister, with the actinobacteria at the base. A major difference between Zehr's parsimony and distance analyses of *nifH* (Zehr et al. 1998, 2000) and the *nifD* phylogenies presented here is the placement of the proteobacterial subgroups. Zehr et al. (1998, 2000) analyses support the monophyly of the proteobacteria as a whole, as well as the monophyly of the α and β subgroups (Zehr et al. 1998, 2000). Our parsimony analysis of amino acid sequences does not support the monophyly of the proteobacteria. However, with the exception of *Acidithiobacillus ferrooxidans*, the monophyly of the proteobacteria is supported by ML analysis of DNA sequences (Fig. 2). None of the analyses presented here resolve the α , β , and γ proteobacterial subgroups as monophyletic. Recent studies suggest that both *nifH* (Udea et al. 1995; Hurek et al. 1997, 1998; Laguerre et al. 2001) and *nifD* (Parker et al. 2002; Qian et al. 2002, 2003) may have been transferred horizontally within the proteobacteria, which would result in failure to differentiate these subgroups. Additionally, it has been suggested that some of the proteobacterial subgroups may not be monophyletic (Garrity and Holt 2001). Although our analyses of *nifD* vary slightly from those of *nifH* (Zehr et al. 1998, 2000), the two are similar in overall topology and support vertical descent. Furthermore, the agreement of *nifH* and *nifD* can be taken as evidence that *nif* genes and nitrogen fixation have evolved by vertical descent.

The analyses of *nifD* presented here are consistent with *nifH* (Zehr et al. 1998, 2000) and resolve the Gram-positive bacteria (firmicutes and actinobacteria), proteobacteria, and cyanobacteria, which is congruent with 16S rRNA phylogenies (Woese 1987; Olsen and Woese 1993; Olsen et al. 1994), thus supporting the vertical descent of *nifD* within these lineages. The discrepancies in earlier *nif* phylogenies may have been due to the number of species that were examined. Previous studies examined a limited number of taxa whereas we examined 57 taxa. These discrepancies could also be due to the methods of analysis used. Previous studies employed only parsimony and/or distance methods. We utilized parsimony, distance, and the model-based maximum likelihood method. It has been suggested that divergent sequences should be analyzed with model-based criteria to more accurately predict phylogeny since noise due to saturation is higher (Swofford et al., 1996; Huelsenbeck and Rannala 1997; Buckley and Cunningham 2002).

Acknowledgments. This research was partially funded by the National Science Foundation–POWRE Grant 9973340 to S.R.B., Miami University Committee on Faculty Research Grant to S.R.B., Ohio Sea Grant Development to S.R.B. and L.E.W., and Ohio Board of Regents Academic Challenge Grant to B.J.H. We also acknowledge the help of Chris Wood and The Center for Functional Genomics and Bioinformatics at Miami University, Eric Stahlberg with the Ohio Super Computer, and Dr. Guillermo Orti for providing advice on the analyses. This paper represents, in part, the senior author's master's thesis.

References

- Bishop PE, Premakumar R (1992) Alternative nitrogen fixation systems. In: Stacey G, Burris RH, Evans HJ (eds.), *Biological Nitrogen Fixation*, Chapman & Hall, New York, pp 736–762
- Borthakur D, Basche M, Buikema WJ, Borthakur P, Haselkorn R (1990) Expression, nucleotide sequence, and mutational analysis of two open reading frames in the *nif* gene region of *Anabaena* sp. strain PCC 7120. *MGG* 221:227–234
- Bremer K (1994) Branch support and tree stability. *Cladistics* 10:295–304
- Duckley TK, Cunningham CW (2002) The effects of nucleotide substitution model assumptions on estimates of nongrammetric bootstrap support. *Mol Biol Evol* 19:394–405
- Fani R, Gallo R, Lio P (2000) Molecular evolution of nitrogen fixation: The evolutionary history of the *nifD*, *nifK*, *nifE*, and *nifN* genes. *J Mol Evol* 51:1–11
- Felsenstein J (1981) Evolutionary trees from DNA sequences: A maximum likelihood approach. *J Mol Evol* 17:368–376
- Felsenstein J (1985) Confidence limits on phylogenies: An approach using the bootstrap. *Evol* 39:783–791
- Garrity GM, Holt JG (2001) The road map to the manual. In: Boone DR, Castenholz RW (eds) *Bergey's manual of systematic bacteriology*, 2nd ed., Vol 1, Springer, New York, pp 119–166
- Haukka K, Lindstrom K, Young JPW (1998) Three phylogenetic groups of *noda* and *nifH* genes in *Sinorhizobium* and *Mesorhizobium* isolates from leguminous trees growing in Africa and Latin America. *Appl Environ Microbiol* 64:419–426
- Haselkorn R, Buikema WJ (1992) Nitrogen fixation in cyanobacteria. In: Stacey G, Burris R, Evans H (eds) *Biological nitrogen fixation*. Chapman & Hall, New York, pp 43–85
- Hennecke H, Kaluzu K, Thony B, Fuhrmann M, Ludwig W, Stackebrandt E (1985) Concurrent evolution of nitrogenase genes and 16S RNA in *Rhizobium* species and other nitrogen fixing bacteria. *Arch Microbiol* 142:342–348
- Hirsch AM, McKhann HI, Reddy A, Liao J, Fang Y, Marshall CR (1995) Assessing horizontal transfer of *nifHDK* genes in Eubacteria: nucleotide sequence of *nifK* from *Frankia* strain HFPCc13. *Mol Biol Evol* 12:16–27
- Huelsenbeck JP, Hillis DM, Jones R (1995) Parametric bootstrapping in molecular phylogenetics: Applications and performance. In: Ferraris JS, Palumbi SR (eds) *Molecular zoology: Strategies and protocols*, Wiley-Liss, New York, pp 19–46
- Huelsenbeck JP, Rannala B (1997) Maximum likelihood estimation of phylogeny using stratigraphic data. *Paleobiology* 23:174–180
- Hurek L, Egener L, Reinhold Hurek B (1997) Divergence in nitrogenases of *Azoarcus* spp., Proteobacteria of the beta subclass. *J Bacteriol* 179:4172–4178
- Kessler PS, McLarnan J, Leigh JA (1997) Nitrogenase phylogeny and the molybdenum dependence of nitrogen fixation in *Methanococcus marisplacidus*. *J Bacteriol* 179:541–543
- Lake JA (1994) Reconstructing evolutionary trees from DNA and protein sequences: Paralinear distances. *Proc Natl Acad Sci USA* 91:1455–1459
- Laguette G, Nour SM, Macheret V, Sanjuan J, Drouin P, Amarger N (2001) Classification of rhizobia based on *nodC* and *nifH* gene analysis reveals a close phylogenetic relationship among *Phaseolus vulgaris*. *Microbiology* 147:981–993
- Lammers PJ, Haselkorn R (1984) Sequence of the *nifD* gene coding for the alpha subunit of dinitrogenase from the cyanobacterium *Anabaena*. *Cell* 44:905–911
- Leigh JA (2001) Nitrogen fixation in methanogens—The Archaeal perspective. In: Triplett EW (ed) *Prokaryotic nitrogen fixation: A model system for analysis of a biological process*. Horizon Scientific Press, Wymondham, UK, pp 657–669
- Lockhart PJ, Steel MA, Hendy MD, Penny D (1994) Recovering evolutionary trees under a more realistic model of sequence evolution. *Mol Biol Evol* 11:605–612
- Maddison DR, Maddison WP (2000) *MacClade* version 4.0. Sinauer Associates, Sunderland, MA
- Mazur BJ, Chui CF (1982) Sequence of the gene coding for the beta-subunit of dinitrogenase from the blue-green alga *Anabaena*. *Proc Natl Acad Sci USA* 79:6782–6786
- Mevarech M, Rice D, Haselkorn R (1980) Nucleotide sequence of a cyanobacterial *nifH* gene coding for nitrogenase reductase. *Proc Natl Acad Sci USA* 77:6476–6480
- Mulligan ME, Haselkorn R (1989) Nitrogen-fixation (*nif*) genes of the cyanobacterium *Anabaena* sp. strain PCC 7120. The *nifB-*fdxN-nifS-nifU** operon. *J Biol Chem* 264:19200–19207
- Mulligan ME, Buikema WJ, Haselkorn R (1988) Bacterial-type ferredoxin genes in the nitrogen fixation regions of the cyanobacterium *Anabaena* sp. strain PCC 7120 and *Rhizobium meliloti*. *J Bacteriol* 170:4406–4410
- Normand P, Bousquet J (1989) Phylogeny of nitrogenase sequences in *Frankia* and other nitrogen-fixing microorganisms. *J Mol Evol* 29:436–444
- Normand P, Gouy M, Cournoyer B, Simonet P (1992) Nucleotide sequence of *nifD* from *Frankia alni* strain Ar13: phylogenetic inferences. *Mol Biol Evol* 9:495–506
- Olsen GJ, Woese CR (1993) Ribosomal RNA: A key to phylogeny. *FSAEB J* 7:113–123
- Olsen GJ, Woese CR, Overbeek R (1994) The winds of (evolutionary) change: Breathing new life into microbiology. *J Bacteriol* 176:1–6
- Parker MA, Lafay B, Burdon JJ, van Berkum P (2002) Conflicting phylogeographic patterns in rRNA and *nifD* indicate regionally restricted gene transfer in *Bradyrhizobium*. *Microbiol* 148:2557–2565
- Posada D, Crandall KA (1998) Modeltest: Testing the model of DNA substitution. *Bioinformatics* 14:817–818
- Postgate JR (1982) *The fundamentals of nitrogen fixation*. Cambridge University Press, New York, pp 1–4
- Postgate JR, Eady RR (1988) The evolution of biological nitrogen fixation. In: Bothe H, de Bruijn F, Newton W (eds) *Nitrogen fixation: Hundred years after*. Proceedings of the 7th international congress on nitrogen fixation. Gustav Fischer, Stuttgart, pp 31–40
- Qian J, Parker MA (2002) Contrasting *nifD* and ribosomal gene relationships among *Mesorhizobium* from *Lotus oroboides* in northern Mexico. *Syst Appl Microbiol* 25:68–73
- Qian J, Kwon SW, Parker MA (2003) rRNA and *nifD* phylogeny of *Bradyrhizobium* from sites across the Pacific Basin. *FEMS Microbiol Lett* 219:13157–13163
- Rice D, Mazur BJ, Haselkorn R (1982) Isolation and physical mapping of nitrogen fixation genes from the cyanobacterium *Anabaena* PCC 7120. *J Biol Chem* 257:13157–13163
- Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406–425
- Sprent J, Sprent P (1990) *Nitrogen fixing organisms: pure and applied aspects*. Chapman and Hall, New York, pp 1–30

- Swofford D (2002) PAUP* 4.0: Phylogenetic analysis using parsimony. Sinauer Associates, Sunderland, MA
- Swofford D, Olsen GJ, Waddell P, Hillis DM (1996) Phylogenetic inference. In: Hillis D, Mortiz C (eds) Molecular systematics. Sinauer Associates, Sunderland, MA, pp 407–513
- Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: Improving the sensitivity of progressive multiple alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680
- Ueda T, Suga Y, Yahiro N, Matsuguchi T (1995) Remarkable N₂-fixing bacterial diversity detected in rice roots by molecular evolutionary analysis of *nifH* gene sequences. *J Bacteriol* 177:1414–1417
- Woese CR (1987) Bacterial evolution. *Microbiol Rev* 51:221–271
- Young JPW (1990) The phylogeny of *nifH*: Gene duplication or lateral transfer? In: Gresshoff PM, Roth LE, Satcey G, Newton WE (eds) Nitrogen fixation: Achievements and objectives. Chapman and Hall, New York, p 840
- Young JPW (1992) Phylogenetic classification of nitrogen-fixing organisms. In: Stacey G, Evans HJ, Burns RH (eds) Biological nitrogen fixation. Chapman and Hall, New York, pp 43–85
- Young JPW (2000) Molecular evolution in diazotrophs: do the genes agree? In: Pedrosa FO, Hungria M, Yates MG (eds) Nitrogen fixation: From molecules to crop productivity. Kluwer, Boston, pp 161–164
- Zehr JP, Mellon MT, Zani S (1998) New nitrogen-fixing microorganisms detected in oligotrophic oceans by amplification of nitrogenase (*nifH*) genes. *Appl Environ Microbiol* 64:3444–3450
- Zehr JP, Carpenter EJ, Villareal TA (2000) New perspectives on nitrogen-fixing microorganisms in tropical and subtropical oceans. *Trends Microbiol* 8:68–73