Phylogeny and Characterization of Three *nifH*-Homologous Genes from *Paenibacillus azotofixans*

Quok-Cheong Choo, Mohd-Razip Samian, and Nazalan Najimudin*

School of Biological Sciences, Universiti Sains Malaysia, 11800 Pulau Pinang, Malaysia

Received 19 November 2002/Accepted 21 March 2003

In this paper, we report the cloning and characterization of three *Paenibacillus azotofixans* DNA regions containing genes involved in nitrogen fixation. Sequencing analysis revealed the presence of *nifBIH1D1K1* gene organization in the 4,607-bp SacI DNA fragment. This is the first report of linkage of a *nifB* open reading frame upstream of the structural *nif* genes. The second (*nifB2H2*) and third (*nifH3*) *nif* homologues are confined within the 6,350-bp *HindIII* and 2,840-bp *EcoRI* DNA fragments, respectively. Phylogenetic analysis demonstrated that NifH1 and NifH2 form a monophyletic group among cyanobacterial NifH proteins. NifH3, on the other hand, clusters among NifH proteins of the highly divergent methanogenic archaea.

Nitrogen fixation-related genes have been highly conserved throughout evolution even though they are widely distributed among eubacteria and archaea (4, 7, 11, 13, 15). In terms of their physical and biochemical properties, the mechanisms of the nitrogen fixation process are very similar among these organisms. The conventional dinitrogenase is composed of an α,β, tetramer; the α and β subunits are encoded by the *nifD* and *nifK* genes, respectively. Also included in the nitrogenase complex is nitrogenase reductase, which is encoded by the *nifH* gene. In most diazotrophs, the *nifHDK* complex is nitrogenase reductase, which is encoded by the *nifH* gene.

Paenibacillus azotofixans ATCC 35681 is a gram-positive, facultatively anaerobic diazotroph that falls into a broad cluster of nitrogen fixers in rRNA group 3; this cluster also includes *P. macerans* and *P. polymyxa* (3). Diazotrophic strains of *P. azotofixans* were shown to possess the ability to fix atmospheric dinitrogen with high efficiency (8, 25, 29). In contrast to the majority of diazotrophs, their ability to fix nitrogen is not affected by the presence of nitrate (29).

**PCR amplification of the *nifH* gene fragment.** The objective of identifying DNA fragments containing *nifH* homologues was achieved by using the 380-bp *nifH*I gene as a homologous probe. Alignment of NifH1 polypeptide sequences from representative diazotrophs was performed using ClustalX software (9). Based on these sequence alignments, *nifH*-degenerate oligonucleotides (5′-TAY GGN AAR GGN GGN ATN GGN AA-3′ and 5′-GCR AAN CCN CCR CAN ACN ACR TC-3′) were designed as primers.

Chromosomal DNA (40 ng/ml) was PCR amplified in a 50-μl reaction volume containing 1× PCR buffer (Promega), a 1 mM concentration of each primer, a 0.2 mM concentration of each deoxynucleoside triphosphate, 1.5 mM MgCl₂, and 2.0 U of Weiss *Taq* DNA polymerase (Promega). The following PCR parameters were used: 94°C for 5 min; 30 thermal cycles of 94°C for 30 s, 45°C for 30 s, and 72°C for 30 s; and a final extension step at 72°C for 10 min.

**Screening of genomic library and Southern blot analysis.** A genomic library of individual lambda clones from primary recombinants was screened according to standard procedures (28), using the PCR-amplified *nifH* probe. Following secondary and tertiary screenings, positively hybridized plaques were isolated and their DNA was extracted. The purified DNA was subjected to restriction enzyme digestions (*EcoRI*, *HindIII*, and SacI). Southern analysis using the *nifH* probe revealed the presence of three distinctly different DNA digestion profiles (data not shown), suggesting the existence of three different *nif* gene-containing DNA regions, which were subsequently gel purified and ligated. Hybridization analysis using the digoxigenin-labeled *nifH* PCR probe was also performed with genomic DNA digested with the *EcoRI* and *HindIII* restriction enzymes (data not shown). The results obtained suggested the presence of more than one copy of the *nifH* gene in *P. azotofixans*, in agreement with previous studies by Oliveira et al. (21) and Rosado et al. (26).

**Sequence analysis of *nifH1*, *nifH2*, *nifH3*, and other *nif* genes.** The 4,607-bp SacI fragment contained a 320-amino-acid partial *nifB*I coding region, the complete *nifH1* and *nifD1* open reading frames (ORFs), and the first 387 amino acids of *nifK1* (Fig. 1a). These alleles were designated *nifB1*, *nifH1*, *nifD1*, and *nifK1*, respectively. This is the first report of linkage of a *nifB* ORF upstream of the conventional structural *nif* genes. Analysis of the region immediately upstream of the *nifH1D1K1* ORFs revealed the presence of potential ribosome binding sites (RBSs; GAGG, GAGG, and GAGG, respectively) (30) located between 8 and 11 bp from the ATG initiation codon of each ORF. The suggested *nifH1* RBS overlaps with the 3′ end of the *nifB1* coding region. Examination of the 143-bp *nifH1-nifD1* intergenic region revealed the presence of an 11-bp inverted-repeat structure that might have a regulatory function during *nifD1K1* transcription. Similar inverted repeats have been described for other diazotrophs (4, 11, 16). Comparison of its amino acid sequences with sequences in the database...
revealed that this nifD1 ORF has the highest degree of homology with members of the gram-positive, high-G+C-content genus Franknia (69% identity with Frankia alni strain ArI3 and 68% identity with Frankia sp. strain EU1K1). A 4-nucleotide overlap occurs between the 3' end of nifD1 and the 5' end of nifK1, an indication of a possible translational coupling phenomenon (22). The sequence of nifK1 was partial from the putative ATG, coding for 387 amino acids, with a putative RBS located 10 bp upstream (within the 3' end of nifD1).

A second nifB-nifH cluster (designated nifB2H2) was found in a 6,350-bp HindIII fragment (Fig. 1b). As with NifH1, the protein coding region of NifH2 is 879 nucleotides in length and encodes a predicted 292-amino-acid polypeptide. Interestingly, as in nifH1, the putative RBS (30) for nifH2 is located within the 5' end of its corresponding nifB2 gene. The nifB2 ORF terminates with a single stop codon, TAA, which is followed by the initiation codon for nifH2 6 bp downstream. Unlike nifB1H1, this nif cluster does not have the nifDK genes within the 3.5-kb region downstream of the nifH2 termination codon. Instead, two potential ORFs that appear to lack any known nif-related function are found. The closest homologies were with various transporter substrate-binding proteins.

The third nifH homologue (nifH3) was found within the 2,840-bp EcoRI fragment (Fig. 1c). No adjacent nifB or nifDK coding regions were found within close proximity. A putative ORF (truncated) that displayed homology to transporter ATP-binding proteins was found approximately 80 bp upstream of the nifH3 start codon.

Amino acid alignment of NifH proteins. Figure 2 shows an alignment of the deduced amino acid sequences of the P. azotofixans NifH proteins. They are more divergent in their C termini. Stretches of 10 or more conserved amino acids were observed for residues 10 to 21, 97 to 109, and 129 to 141. When the amino acid residues of NifH1 and NifH2 were compared, seven were found to differ; this constitutes 97% identity. Comparing either NifH1 or NifH2 with NifH3 yielded a comparatively low 43% identity. A high (97%) identity was also observed when the partial reading frames of nifB genes were translated to their respective amino acids. At the nucleotide level, alignment of the two nifBH gene clusters also revealed a high degree of identity (94%), with no significant changes until 57 bases downstream of the presumptive termination codon of the nifH ORF (data not shown). These data led us to postulate that nifBH gene clusters of P. azotofixans had undergone a gene duplication process, resulting in the nifB1H1 and nifB2H2 gene organizations.

Phylogenetic analysis. The nifH phylogenetic tree had been well established (20, 26, 32, 35, 36) and is largely consistent with the 16S rRNA gene phylogeny (34). Our data (Fig. 3) are in agreement with the division of the NifH topology into four major clusters, as described by Chien and Zinder (5, 6). When the complete nifH coding sequences were used, the clustering of P. azotofixans NifH1 and NifH2 yielded several interesting observations. Earlier nifH-based phylogenetic analyses of P. azotofixans involved partial sequences of nifH fragments derived by PCR amplification (1, 26, 36). Discrepancies between our study and those of other investigators (1, 26, 36) in the placement of P. azotofixans NifH proteins were probably due to their use of short-length nifH fragments, which reduced the resolving power of the analyses. When phylogeny was based on partial nifH gene sequences, determined by Zehr et al. (36), P. azotofixans NifH did not cluster with NifH proteins of any group of bacteria. Further observations and the branching order of the NifH phylogeny seemed to suggest that P. azotofixans NifH lies within the cyanobacterial clade (1, 36).

Use of the complete DNA sequences of the three nifH genes in a reanalysis of NifH phylogeny demonstrated clustering of P. azotofixans NifH1 and NifH2 within the Cyanobacteriaceae grouping (Fig. 3). The NifH protein from a filamentous, non-heterocystous marine cyanobacterium, Trichodesmium sp. strain IMS101, showed the highest degrees of identity with P. azotofixans NifH1 (80%) and NifH2 (79%), respectively. Interestingly, neither NifH1 nor NifH2 clustered with the NifH proteins of other gram-positive diazotrophs, such as Frankia spp. (a high-G+C firmicute) and Clostridium pasteurianum (a low-G+C firmicute).

The third putative nifH gene product of P. azotofixans (NifH3) clustered with NifH proteins of members of the Archaean domain, Methanothermococcus thermolithothrophicus and Methanothermobacter thermoautotrophicus. Again, this putative NifH did not cluster with those of the other phylogenically related gram-positive microorganisms, such as Frankia spp. or C. pasteurianum. This is the first report of a gram-positive diazotroph
having a NifH protein that clusters with Nif proteins of confirmed methanogenic diazotrophs. Based on the NifH phylogenetic analysis (Fig. 3), P. azotoxans NifH3 also did not fall within the anf nitrogenase clade. Rosado et al. (26) reported a nifH phylogenetic tree in which the proteins of three Paenibacillus strains, P. azotoxans P3E20 and RBN4 and P. durum DSMZ1735, formed a cluster with the alternative (anf) nitrogenases. It is not known at this point whether the putative anf nitrogenase reported by Rosado et al. (26) exists in this P. azotoxans type strain as well. It has not yet been determined whether this nifH3 gene product is a functional nitrogenase. It has been postulated that genes from this cluster are related to genes involved in bacteriochlorophyll synthesis and probably have a function unrelated to nitrogen fixation (5, 12).

The question of horizontal transfer of the nifH gene has been debated among evolutionists for the last 3 decades (10, 17, 18). The strongest evidence yet for horizontal nifH gene transfer came from the pioneering phylogeny studies of nifH genes (17, 18). Verification of a horizontal-transfer event is difficult, especially with the limited genetic data from Paenibacillus strains. Nevertheless, our data on NifH phylogeny revealed some unanticipated features that brought us to postulate that the gene transfer phenomenon exists. The most striking evidence for the occurrence of a gene transfer event was the unusual placement of NifH3 among the highly divergent members of the Archaea. Smith et al. (31) described a phylogenetic congruency test based on the assumption that a NifH tree corresponds to conventional NifH phylogenies (20, 26, 32, 35, 36); if there was any odd placement, a horizontal-gene-transfer event may have occurred. Furthermore, the low level of identity (43%) between P. azotoxans NifH3 and the other two NifH proteins likely indicates that there are two different groupings of orthologous gene products. The vast differences in the sequences among NifH3 proteins compared to NifH1 and NifH2 seemed to suggest that a duplication event was unlikely; otherwise, like the five C. pasteurianum NifH proteins (33), all three P. azotoxans NifH proteins would be grouped in the same cluster.

It is not presently known whether all three nif homologue clusters are located in the genome or on plasmids (if any even...
exist), as in some diazotrophic systems (19, 24). It will also be of interest to determine whether the phylogenies of complete \textit{nifH} genes of other \textit{Paenibacillus} strains conform to the conventional \textit{nifH} phylogenetic topology.

**Nucleotide sequence accession numbers.** The sequencing data obtained in this study have been deposited in the EMBL database under the following accession numbers: AJ299453, AJ299454, and AJ515294.
This work was supported by research grants from the Malaysian Ministry of Science, Technology and Environment (MOSTE); a Universiti Sains Malaysia short-term grant; and the Malaysian Toray Science Foundation (MTSFS).

REFERENCES


