Molecular evolution of the *nif* gene cluster carrying *nifl*₁ and *nifl*₂ genes in the Gram-positive phototrophic bacterium *Heliobacterium chlorum*

Jigjiddorj Enkh-Amgalan, Hiroko Kawasaki and Tatsuji Seki

Correspondence

Hiroko Kawasaki
ICBKawasakiNakagawa@icb.
osaka-u.ac.jp

The International Center for Biotechnology, Osaka University, 2-1 Yamada-oka, Suita-shi, Osaka 565-0871, Japan

A major *nif* cluster was detected in the strictly anaerobic, Gram-positive phototrophic bacterium *Heliobacterium chlorum*. The cluster consisted of 11 genes arranged within a 10 kb region in the order *nifl*₁, *nifl*₂, *nifH*, *nifD*, *nifK*, *nifE*, *nifN*, *nifX*, *fdx*, *nifB* and *nifV*. The phylogenetic position of *Hbt. chlorum* was the same in the NifH, NifD, NifK, NifE and NifN trees; *Hbt. chlorum* formed a cluster with *Desulfitobacterium hafniense*, the closest neighbour of heliobacteria based on the 16S rRNA phylogeny, and two species of the genus *Geobacter* belonging to the *Deltaproteobacteria*. Two *nifl* genes, known to occur in the *nif* clusters of methanogenic archaea between *nifH* and *nifD*, were found upstream of the *nifH* gene of *Hbt. chlorum*. The organization of the *nif* operon and the phylogeny of individual and concatenated gene products showed that the *Hbt. chlorum* *nif* operon carrying *nifl* genes upstream of the *nifH* gene was an intermediate between the *nif* operon with *nifl* downstream of *nifH* (group II and III of the nitrogenase classification) and the *nif* operon lacking *nifl* (group I). Thus, the phylogenetic position of *Hbt. chlorum* nitrogenase may reflect an evolutionary stage of a divergence of the two nitrogenase groups, with group I consisting of the aerobic diazotrophs and group II consisting of strictly anaerobic prokaryotes.

INTRODUCTION

The description of *Heliobacterium* (*Hbt.*) *chlorum* in 1983 led to the establishment of the new genus *Heliobacterium* and added a new member to the list of groups of photosynthetic organisms (Gest & Favinger, 1983). This strictly anaerobic organism contained a novel photosynthetic pigment, bacteriochlorophyll *g*, which became a main characteristic of the family '*Heliobacteriaceae*' (Brockmann & Lipinski, 1983; Madigan, 2001). Phylogenetically, heliobacteria belong to the group of low-G + C-content Gram-positive bacteria that includes *Clostridium* and *Bacillus* (Madigan, 1992, 2001), and they all have the capacity for endospore formation (Kimble-Long & Madigan, 2001). *Hbt. chlorum* was isolated from garden soil (Gest & Favinger, 1983), and later a large number of heliobacteria was isolated from rice fields (Beer-Romero & Gest, 1987; Ormerod *et al.*, 1996; Stevenson *et al.*, 1997), hot springs (Kimble *et al.*, 1995; Stevenson *et al.*, 1997) and the banks of soda lakes (Bryantseva *et al.*, 1999, 2000); the main source of

heliobacteria seems to be rice fields (Stevenson *et al.*, 1997; Madigan, 2001). The link between the heliobacterial habitat (rice field) and the ability of heliobacteria to fix nitrogen both photosynthetically and in darkness suggested that they might be significant contributors of fixed nitrogen in the rice fields (Kimble & Madigan, 1992).

Nitrogen fixation is widely but sporadically distributed among both eubacteria and methanogenic archaea (Young, 1992; Raymond *et al.*, 2004). The current understanding of nitrogenase diversity has been based largely on phylogenetic analyses of *nifH* and *nifD*, the nitrogenase structural genes (Zehr *et al.*, 2003; Henson *et al.*, 2004). Recently, Raymond *et al.* (2004) performed genomic analyses of *nif* genes encoding the core components of nitrogenase, including the NifH, NifD, NifK, NifE and NifN proteins, and proposed five groups: (1) typical Mo–Fe nitrogenases, predominantly composed of members of the proteobacterial and cyanobacterial phyla; (2) anaerobic Mo–Fe nitrogenases from predominantly anaerobic bacteria and several methanogens; (3) alternative nitrogenases, including the Mo-independent Anf and Vnf proteins (except VnfH, which is more similar to NifH rather than AnfH); (4) uncharacterized *nif* homologues detected only in methanogens and some anoxygenic photosynthetic bacteria; and (5) bacteriochlorophyll and chlorophyll biosynthesis genes common to all phototrophs.

Published online ahead of print on 2 September 2005 as DOI 10.1099/ij.s.0.63815-0.

The GenBank/EMBL/DDBJ accession number for the complete coding sequence of the *nif* gene cluster of *Hbt. chlorum* DSM 3682^T is AB196525.

This grouping was largely consistent with the previous classification, in which the nitrogenase genes were divided into clusters I–IV (Zehr *et al.*, 2003).

Phylogenetic analyses of NifH and NifD sequences of heliobacteria showed that heliobacteria form a distinct lineage in the nitrogenase phylogeny. Although heliobacteria are strictly anaerobic bacteria, they did not belong to group II of strictly anaerobic diazotrophs such as *Clostridium*, and were instead placed in group I (Enkh-Amgalan *et al.*, 2005). Indeed, the approximately 50-residue conserved insertion in *nifD* shared by all members of group II was not found in heliobacteria. Moreover, the *nifH* and *nifD* genes of heliobacteria were contiguous, unlike *nifH* and *nifD* of members of group II, which are separated by two *glnB*-like genes (Chien & Zinder, 1996; Kessler *et al.*, 1998; Kessler & Leigh, 1999; Sibold *et al.*, 1991; Arcondeguy *et al.*, 2001; Chen *et al.*, 2001; unpublished genome survey) recently designated *nifl*₁ and *nifl*₂ (Arcondeguy *et al.*, 2001). Another interesting finding was the specific relationship of heliobacteria with *Geobacter* species, which belong to the *Deltaproteobacteria*; the clade of heliobacteria was grouped with the *Geobacter* species clade in the NifH phylogeny, whereas no such grouping was formed in the NifD phylogeny, and both heliobacteria and *Geobacter* species formed independent clades (Enkh-Amgalan *et al.*, 2005).

Consequently, we aimed to isolate and analyse other genes involved in nitrogen fixation of strictly anaerobic heliobacteria in order to understand fully their unique position in group I and to search for genes common to group I and/or group II diazotrophs. We selected *Hbt. chlorum*, the type species of *Heliobacterium*, for this study. Interestingly, sequencing results revealed similar features between the *Hbt. chlorum* *nif* cluster and group II of strictly anaerobic diazotrophs, i.e. small size and the presence of *nifl* genes (found upstream of *nifH*). The gene organization and phylogenetic analyses of the *nifl* and other *nif* genes of *Hbt. chlorum* in comparison with those of other diazotrophs are discussed in order to further understanding of nitrogenase evolution.

METHODS

Bacterial strain and DNA preparation. *Hbt. chlorum* DSM 3682^T was grown phototrophically (anoxic/light) in completely filled 50-ml screw-capped bottles with DSM medium 370 (<http://www.dsmz.de>) at 37 °C for 2 days. *Escherichia coli* strains for recombinant DNA manipulations were grown in Luria–Bertani medium at 37 °C. Genomic DNA was isolated as described by Ausubel *et al.* (1995).

DNA manipulations and cloning procedures. All DNA manipulations were performed using standard techniques (Sambrook *et al.*, 1989) and according to instructions provided by the suppliers of the reagents. Chromosomal DNA was digested with appropriate restriction enzymes (TaKaRa Shuzo), fractionated by electrophoresis on 0.7% agarose gel and transferred to Hybond-N+ nylon membrane (Amersham Biosciences) by capillary transfer. Hybridization probes from the PCR product were labelled using the DIG labelling kit (Roche Diagnostics), and hybridization signals were detected using the DIG luminescence detection kit (Roche Diagnostics) according

Fig. 1. *Hbt. chlorum* nitrogen-fixation gene cluster. Arrows represent cloned DNA fragments and thick lines represent probes used in Southern analysis. B, *Bam*HI; E, *Eco*RI; K, *Kpn*I; P, *Pst*I.

to the manufacturer's instructions. We used four probes and the positions of the probes are shown in Fig. 1. DNA fragments that hybridized to the probes were recovered from agarose gel using the QIAEX II gel extraction kit (Qiagen) and cloned into pUC18 using the DNA ligation kit version 1 (TaKaRa Shuzo). Recombinant colonies were transferred onto Hybond-N+ nylon membrane (Amersham Biosciences), and hybridization and detection were performed as in Southern hybridization. Plasmid DNAs were purified using the QIAprep Spin Miniprep kit (Qiagen) for sequencing analysis.

Sequencing and sequence analysis. DNA sequences were determined by the dideoxy chain-termination method (Sanger *et al.*, 1977) with a BigDye Terminator cycle sequencing kit (PE Applied Biosystems) using an ABI PRISM 310 Genetic Analyzer (Perkin Elmer). Sequence data were analysed by the ABI PRISM sequence analysis program and assembled using the ABI Auto Assembler (Perkin Elmer). Nucleotide sequences were analysed with GENETYX-WIN software (version 3.1). Sequence similarity searches were performed via BLAST (Altschul *et al.*, 1997) at both the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>) and the DNA Database of Japan (DDBJ; <http://www.ddbj.nig.ac.jp/>). Inferred amino acid sequences of individual genes were aligned with those of other bacteria and archaea by the CLUSTAL X program (Thompson *et al.*, 1997; Jeanmougin *et al.*, 1998) and checked by hand for proper alignment. Phylogenetic analysis was performed using the CLUSTAL X program. Alignment positions in which any of the sequences had a gap were discarded by using the 'exclude positions with gaps' option, and evolutionary distances were corrected for multiple substitutions by using the appropriate option in the settings. The protein weight matrix GONNET 250 was used for sequence comparison, and phylogenetic trees were constructed by the neighbour-joining method (NJ; Saitou & Nei, 1987) with 1000 bootstrap replicates using default parameters. The concatenated NifHDKEN and Nifl trees were degenerated by both the NJ and maximum-likelihood (ML) methods. The programs ProML, SeqBoot and CONSENSE from PHYLIP (Felsenstein, 2004) were used to infer and assemble ML trees (100 replicates), using the JTT model of amino acid substitution. To display and analyse the tree, NJPlot (Perrière & Gouy, 1996) and TreeView (Page, 1996) were used.

RESULTS AND DISCUSSION

Isolation and identification of nitrogen-fixation genes

Previously, we showed that the *nifH* and *nifD* genes of *Hbt. chlorum* are contiguous and that *nifH* is probably present in

a single copy. Hybridization with a *Heliobacterium gestii* *anfH* probe suggested that no alternative nitrogenase is present in *Hbt. chlorum*. *Bam*HI digests of *Hbt. chlorum* genomic DNA contained a 2·4 kb fragment homologous to the *nifH* probe (Enkh-Amgalan *et al.*, 2005). In this study, *Hbt. chlorum* genomic DNA was digested with *Bam*HI and approximately 2·4 kb fragments were isolated and ligated into the *Bam*HI site of pUC18. Colony hybridization resulted in five colonies giving a strong positive signal to the *nifH* probe. Plasmid DNA from the positive colonies was purified and then sequenced by primer walking and shown to carry the complete *nifH* gene and part of the *nifD* gene (Fig. 1). Based on the sequence information obtained, specific probes were designed and the flanking regions were cloned and sequenced by chromosome walking until regions of genes not involved in nitrogen fixation were obtained. The probes used and the size and coding region of the fragments obtained are shown in Fig. 1. In this way, we obtained overlapping fragments with a total nucleotide sequence of 11 234 bp. Analysis of the entire nucleotide sequence revealed the presence of 11 open reading frames (ORFs). The amino acid sequences deduced from the nucleotide sequence of the ORFs showed remarkable similarity to the products of nitrogen-fixation genes in other diazotrophs (Table 1). The highest level of identity was detected with predicted gene products derived from a genomic sequence of *Desulfitobacterium hafniense*. Indeed, *Desulfitobacterium* is the genus that is most closely related to heliobacteria based on the 16S rRNA phylogeny, and these bacteria share some other similarities such as the formation of endospores and the absence of an outer membrane, despite negative Gram staining (Niggemeyer *et al.*, 2001). However, it should be noted that the capacity for nitrogen fixation has not yet been detected in the genus *Desulfitobacterium*.

Nitrogen-fixation gene cluster

The *Hbt. chlorum nifH* gene was encoded by an ORF of 855 bp (including the stop codon) and the ORF corresponding to *nifD* started 70 bp downstream from the termination codon of *nifH*. The number of amino acids and the product sizes are shown in Table 1. A 1 bp overlap was observed between the 3' end of *nifD* and the 5' end of the next ORF, *nifK*, which indicates a possible translational coupling phenomenon (Oppenheim & Yanofsky, 1980). *nifH*, *nifD* and *nifK* started with the same start codon, ATG, and terminated with the stop codon TAA. The *nifE* gene, 109 bp downstream of *nifK*, had a less commonly used translational initiation codon, GTG, which is also found in some *nif* genes in various diazotrophs, for example, *nifB* and *nifX* of *Frankia alni* (Harriott *et al.*, 1995) and *nifN* of *Acetobacter diazotrophicus* (Lee *et al.*, 2000). Adjacent to the *nifE* gene (separated by 40 bp) was *nifN*, which started with the initiation codon ATG and terminated with the stop codon TAA. Analysis of the region immediately upstream of these five genes revealed the presence of putative ribosome-binding sites (AGGAGG, AAGAGG and AGGAGG) respectively located 8–10 bp from the start codons of *nifH*, *nifD*

Table 1. Comparison of the *nif* products of *Hbt. chlorum* with the equivalent proteins in other diazotrophs, and possible functions

Genus abbreviations: *A.*, *Azotobacter*; *An.*, *Anabaena*; *D.*, *Desulfitobacterium*; *G.*, *Geobacter*; *Mb.*, *Methanobacterium*; *Mc.*, *Methanococcus*; *Ms.*, *Methanosarcina*; *P.*, *Paenibacillus*; *Rps.*, *Rhodopseudomonas*. Amino acid sequence identity and numbers of positives were obtained from BLAST search results.

Gene	Product		Organisms with most similar gene products (% amino acid identity; positives)	Function
	Amino acids (<i>n</i>)	Molecular mass (kDa)		
<i>nifH</i> ₁	105	11·6	<i>D. hafniense</i> (74; 90), <i>Ms. mazei</i> (62; 74), <i>Ms. acetivorans</i> (60; 73)	Regulation of nitrogenase activity; 'switch-off'
<i>nifH</i> ₂	127	13·9	<i>D. hafniense</i> (71; 80), <i>Mc. maripaludis</i> (50; 63), <i>Mb. ivanovii</i> (48; 67)	Regulation of nitrogenase activity; 'switch-off'
<i>nifH</i>	284	31·0	<i>D. hafniense</i> (82; 91), <i>G. metallireducens</i> (83; 92), <i>A. vinlandii</i> (77; 89)	Nitrogenase structure; Fe protein
<i>nifD</i>	488	54·9	<i>D. hafniense</i> (74; 87), <i>G. sulfurreducens</i> (67; 82), <i>P. durus</i> (67; 81)	Nitrogenase structure; MoFe protein α subunit
<i>nifK</i>	480	52·8	<i>D. hafniense</i> (62; 78), <i>G. sulfurreducens</i> (57; 72), <i>An. variabilis</i> (56; 75)	Nitrogenase structure; MoFe protein β subunit
<i>nifE</i>	453	49·0	<i>D. hafniense</i> (71; 85), <i>G. sulfurreducens</i> (61; 78), <i>Nostoc</i> sp. (51; 71)	FeMo co-factor biosynthesis
<i>nifN</i>	432	46·7	<i>D. hafniense</i> (65; 76), <i>G. metallireducens</i> (53; 68), <i>An. variabilis</i> (40; 60)	FeMo co-factor biosynthesis
<i>nifX</i>	123	13·5	<i>D. hafniense</i> (50; 62), <i>Rps. palustris</i> (43; 65), <i>G. sulfurreducens</i> (40; 61)	FeMo co-factor biosynthesis
<i>fix</i>	88	9·7	<i>D. hafniense</i> (50; 70), <i>G. sulfurreducens</i> (52; 65), <i>Nostoc</i> sp. (51; 63)	Ferredoxin-like protein
<i>nifB</i>	277	30·0	<i>D. hafniense</i> (58; 75), <i>G. sulfurreducens</i> (55; 72), <i>Ms. acetivorans</i> (51; 71)	FeMo co-factor biosynthesis
<i>nifV</i>	375	41·4	<i>D. hafniense</i> (52; 69), <i>G. sulfurreducens</i> (46; 64), <i>Rps. palustris</i> (47; 62)	FeMo co-factor biosynthesis; homocitrate synthase

and *nifE*. In the region flanking the 3' end of *nifK*, an inverted-repeat sequence that might function as a transcription terminator was present 40 bp downstream from *nifK*. Such inverted repeats have been found in the *nif* operons of other diazotrophs and are suggested to have a transcriptional regulatory function (Brigle *et al.*, 1985; Norel & Elmerich, 1987; Minerdi *et al.*, 2001). However, the inverted-repeat structure was not found in the *nifH*–*nifD* intergenic region, unlike in the other diazotrophs. Downstream of the *nifHDKEN* genes were the following genes in this order: *nifX*, *fdx*, *nifB* and *nifV*. *nifN* and *nifX* were separated by 135 bp, 12 bp separated *nifX* and *fdx*, 170 bp separated *fdx* and *nifB* and *nifV* was 17 bp downstream of *nifB*. The inverted-repeat sequence was detected 105 bp downstream from the stop codon of *nifV* only.

Upstream of *nifH*, we detected two small ORFs encoding 105 and 127 amino acids. The products of these two ORFs exhibited significant similarity to those of the *nifI*₁ and *nifI*₂ genes present in the *nif* operon of methanogenic archaea (Chien & Zinder, 1996; Kessler *et al.*, 1998; Kessler & Leigh, 1999; Sibold *et al.*, 1991) and in some strictly anaerobic bacteria, *Desulfovibrio gigas*, *Clostridium acetobutylicum*, *Clostridium cellobioparum* (Arcondeguy *et al.*, 2001), *Clostridium beijerinckii* (Chen *et al.*, 2001), *Desulfovibrio vulgaris* and *Chlorobium tepidum* (unpublished genome survey), which belong to group II of nitrogenase. In all cases, *nifI*₁ and *nifI*₂ are located between *nifH* and *nifD*, suggesting their conserved function (Arcondeguy *et al.*, 2001). However, in the previously demonstrated NifH- and NifD-based phylogeny, sequences of the strictly anaerobic heliobacteria were placed in nitrogenase group I, indicating that they bore less similarity to group II, which consists of strictly anaerobic diazotroph sequences (Enkh-Amgalan *et al.*, 2005). Thus, the finding of *nifI* genes that exhibit striking similarity to the *nifI* genes of members of group II in both sequence identity and product size in heliobacteria was unexpected. Furthermore, in the *Hbt. chlorum* *nif* operon, *nifI* genes were found upstream of *nifH*; this is the first report of *nifI* genes located upstream of the nitrogenase structural genes. A BLAST search of the *nifI*₁ and *nifI*₂ amino acid sequences against the genome sequence of *D. hafniense* resulted in 74 and 71 % identical products, respectively, that were located upstream of *nifH*. The observation of *nifI* genes in the same location in the *nif* clusters of *Hbt. chlorum* and *D. hafniense*, together with the high identity of individual *nif* gene products and the similar overall organization of genes within the *nif* cluster in both organisms, suggested that these bacteria have remarkably similar nitrogenase systems.

Phylogenetic analysis

To study the evolutionary relationships between *Hbt. chlorum* and other nitrogen-fixing prokaryotes, products of *nifH*, *nifD*, *nifK*, *nifE* and *nifN* genes which encode the core components of nitrogenase were compared with corresponding sequences in the DDBJ/EMBL/GenBank databases and phylogenetic trees were generated (Fig. 2). Overall topologies of the five trees were significantly consistent with

each other and, in particular, the phylogenetic position of *Hbt. chlorum* among other diazotrophs was consistently preserved in all trees. In the NifH, NifK, NifE and NifN trees, the sequences of *Hbt. chlorum* formed a cluster with sequences from *D. hafniense* and two metal-reducing bacteria in the *Deltaproteobacteria*, *Geobacter sulfurreducens* and *Geobacter metallireducens*, and the cluster was placed in group I, as expected. However, in the NifD tree, the two *Geobacter* species formed an independent cluster (with a low bootstrap value) which branched earlier than the cluster of *Hbt. chlorum* and *D. hafniense*. Nevertheless, these two clusters were placed as the deepest lineages in group I. Actually, *D. hafniense* was shown to bridge the gap between the group I and group II clades in the NifD, NifK, NifE and NifN trees, while NifH was found within group I but with poor bootstrap support (Raymond *et al.*, 2004), which was probably because of the data available for the analysis.

In order to understand fully the phylogenetic history of the nitrogenase system, we concatenated sequences of NifH, NifD, NifK, NifE and NifN and constructed phylogenetic trees using the NJ (left) and ML (right) methods (Fig. 3). Alternative nitrogenases or Anf (lack EN paralogues) and Vnf (VnfH clusters with NifH while VnfDK forms a separate clade) proteins and that of *Rhodospseudomonas palustris*, which is thought to have acquired the *nifH* gene by lateral transfer (Cantera *et al.*, 2004), were excluded from this analysis. As shown in Fig. 3, the NJ and ML trees were consistent with each other and the phylogenetic relationships of diazotrophs were supported by the finding of higher bootstrap values than in the case of individual proteins. Sequences of *Hbt. chlorum*, *D. hafniense*, *G. sulfurreducens* and *G. metallireducens* formed a highly supported clade whose position was preserved as the basal lineage in group I. These data supported the phylogenetic position of *Hbt. chlorum* on the individual protein trees.

The positions of the *nifHDKEN* genes within the *nif* operon in each diazotroph (including the *nifI* genes in organisms that carry them) were also compared, and the comparison demonstrated an interesting picture of *nifI* gene evolution. The arrangement of the *Hbt. chlorum* and *D. hafniense* *nif* operons, carrying *nifI* genes upstream of the *nifH* gene, was an intermediate between the *nif* operon with *nifI* downstream of *nifH* (group II) and the *nif* operon lacking *nifI* (group I). The roles of *nifI* have been studied only in *Methanococcus maripaludis*, and this study demonstrated that both genes are required for the switch-off type of nitrogenase regulation. However, the switch-off mechanism is probably novel, since the covalent modification of dinitrogenase reductase, which occurs during the bacterial switch-off, was not detected (Kessler *et al.*, 2001). Interestingly, the *G. sulfurreducens* and *G. metallireducens* *nif* operons lacked the *nifI* genes, but the *draT* and *draG* genes, which are involved in the bacterial switch-off, were found.

NifI proteins belong to the P_{II} signal transduction protein family, which consists of a large number of GlnB, GlnK and NifI proteins (Ninfa & Atkinson, 2000; Arcondeguy *et al.*,

Fig. 2. Phylogenetic trees for NifH, NifD, NifK, NifE and NifN showing the phylogenetic position of *Hbt. chlorum* among all known corresponding sequences of various diazotrophs. Trees were constructed by the neighbour-joining method and bootstrap values above 50% from 1000 resamplings are shown for each node. Sequences obtained in this study and branches supported by bootstrap values of 90% or more are shown in bold. Bacterial name codes and GenBank accession numbers of each protein are shown in Table 2.

2001). We explored the phylogeny based on NifI together with GlnB and GlnK sequences using the NJ and ML methods (Fig. 4). Both phylogenetic trees were divided into

three subfamilies, NifI₁, NifI₂ and GlnB with GlnK, and this division was in agreement with phylogenetic trees drawn by other authors (Chien & Zinder, 1996; Noda *et al.*, 1999). The

Table 2. Strains and GenBank/DDBJ/EMBL accession/protein ID numbers of *nifHDKEN* sequences used in this study

Species/strain	Code	Accession/protein ID number				
		<i>nifH</i>	<i>nifD</i>	<i>nifK</i>	<i>nifE</i>	<i>nifN</i>
<i>Rhizobium etli</i> CFN 42 ^T	Retli	U80928	U80928	U80928	U80928	U80928
<i>Mesorhizobium loti</i> MAFF 303099	Mloti	AP003005	AP003005	AP003005	AP003005	AP003005
<i>Sinorhizobium meliloti</i> 1021	Smeli	AE007235	AE007235	AE007236	AE007236	AE007238
<i>Gluconacetobacter diazotrophicus</i> PA1 5 ^T	Gacd	AF030414	AF030414	AF030414	AF030414	AF030414
<i>Rhodobacter capsulatus</i>	nRbcap	M15270	M15270		X17433	X17433
<i>Rhodobacter capsulatus</i> B10S	aRbcap	X70033	X70033	X70033		
<i>Rhodopseudomonas palustris</i> CGA009	nRppal	BX572607	BX572607	BX572607	BX572607	BX572607
<i>Rhodopseudomonas palustris</i> CGA009	aRppal	BX572597	BX572597	BX572597		
<i>Rhodopseudomonas palustris</i> CGA009	vRppal	BX572597	BX572597	BX572597	BX572597	BX572597
<i>Rhodobacter sphaeroides</i> ATH 2.4.1 ^T	Rbsph	ZP_00007624	ZP_00007625	ZP_00007626	ZP_00207062	ZP_00207063
<i>Rhodospirillum rubrum</i>	Rsrub	ZP_00269733	ZP_00269732	ZP_00269731	ZP_00267748	ZP_00267747
<i>Azospirillum brasilense</i> ATCC 29145 ^T	Asbra	M64344	M64344	M64344	AF361867	AF361867
<i>Herbaspirillum seropedicae</i> Z78	Hsser	Z54207	Z54207	Z54207	AF088132	AF088132
<i>Burkholderia fungorum</i> LB400	Bfun	ZP_00282264	ZP_00282265	ZP_00282266	ZP_00282269	ZP_00282270
<i>Azoarcus</i> sp. BH72	Aasp	AF200742	AF200742	AF200742		
<i>Nostoc</i> sp. PCC 7120	Noss	BAB73411	BAB73398	BAB73397	BAB73395	BAB73394
<i>Cyanotheca</i> sp. PCC 8801	Cthsp	U22146	U22146	U22146	AF003700	AF003700
<i>Trichodesmium erythraeum</i> IMS101	Tder	ZP_00327022	ZP_00327023	ZP_00327024	ZP_00327025	ZP_00327025
<i>Plectonema boryanum</i> M101	Pbor	D00666	D00666			
<i>Klebsiella pneumoniae</i>	Kp	X13303	X13303	X13303	X13303	X13303
<i>Azotobacter vinelandii</i> OP	nAbvin	M20568	M20568	M20568	M20568	M20568
<i>Azotobacter vinelandii</i> CA	aAbvin	M23528	M23528	M23528		
<i>Azotobacter vinelandii</i> CA	vAbvin	M32371	M32371	M32371		
<i>Thiobacillus ferrooxidans</i> ATCC 33020	Tfer	M15238	M15238	M15238		
<i>Pseudomonas stutzeri</i> A1501	Pms	Q44044	Q44045	AJ313205	AJ313205	
<i>Magnetococcus</i> sp. MC-1	Mcsp	ZP_00291144	ZP_00291143	ZP_00291142	ZP_00291137	ZP_00291136
<i>Paenibacillus azotofixans</i> ATCC 35681 ^T	Pbaz	AJ515294	AJ515294			
<i>Frankia</i> sp. EuK1	Frsp	U53362	U53362	U53362	AF119361	AF119361
<i>Geobacter sulfurreducens</i> PCA ^T	Gbsul	AE017217-77	AE017217-76	AE017217-75	AE017217-62	AE017217-62
<i>Geobacter metallireducens</i> GS-15 ^T	Gbm	ZP_00300753	ZP_00300754	ZP_00300755	ZP_00300759	ZP_00300759
<i>Desulfotobacterium hafniense</i> DCB-2 ^T	Dbh	ZP_00099588	ZP_00099589	ZP_00099590	ZP_00099591	ZP_00099592
<i>Desulfovibrio vulgaris</i> Hildenborough	Dvv	AE017286	AE017286	AE017286	AE017286	AE017286
<i>Chlorobium tepidum</i> TLS ^T	Chtep	AE012909	AE012909	AE012909	AE012909	AE012909
<i>Clostridium pasteurianum</i> W5	Clp	AY603957	AY603957	AY603957	AY603957	AY603957
<i>Clostridium acetobutylicum</i> ATCC 824 ^T	Cla	AE007538	AE007538	AE007538	AE007539	AE007539
<i>Clostridium beijerinckii</i> NRRL B-593	Clb	AF266462	AF266462	AF266462	AF266462	AF266462
<i>Methanococcus maripaludis</i> LL	Mcm	U75887	U75887	U75887	U75887	U75887
<i>Methanobacterium thermoautotrophicum</i> ΔH ^T	Mbt	AE000916	AE000916	AE000916	AE000916	AE000916
<i>Methanobacterium ivanovii</i>	Mbi	X56071				
<i>Methanosarcina acetivorans</i> C2A ^T	nMsa	AE011101	AE011101	AE011101	AE011102	AE011102
<i>Methanosarcina acetivorans</i> C2A ^T	aMsa	AE010788	AE010788	AE010788		
<i>Methanosarcina acetivorans</i> C2A ^T	vMsa	AE010789	AE010789	AE010789		
<i>Methanosarcina barkeri</i> fusaro	nMsb	ZP_00295507	ZP_00295504	ZP_00295503	ZP_00295502	ZP_00295501
<i>Methanosarcina barkeri</i> fusaro	vMsb	ZP_00295459	ZP_00295456	ZP_00295454		
<i>Methanosarcina barkeri</i> fusaro	aMsb		ZP_00297291	ZP_00297289		
<i>Methanosarcina mazei</i> Go1	Msm	AE013297	AE013297	AE013297	AE013298	AE013298
<i>Leptospirillum ferrooxidans</i> L3.2	Lsfer	AF547999	AF547999			
<i>Hellobacterium chlorum</i> DSM 3682 ^T	Hbtc	AB196525	AB196525	AB196525	AB196525	AB196525

Hbt. chlorum Nifl₁ and Nifl₂ sequences fell into the Nifl₁ and Nifl₂ subfamilies, respectively. Within each subfamily, the *Hbt. chlorum* Nifl sequences formed a clade with that of *D.*

hafniense, and this clade was distinct from clades formed by members of nitrogenase group I and group II. The presence of *nifl* genes in both the *Archaea* and *Bacteria* suggests the

Fig. 3. Phylogenetic trees for concatenated NifHDKEN sequences created with the NJ and ML methods. Bootstrap values above 50 % are shown for each node. Sequences obtained in this study and branches supported at 90 % or more are shown in bold. Genus abbreviations: A., *Azotobacter*; As., *Azospirillum*; B., *Burkholderia*; C., *Clostridium*; Chl., *Chlorobium*; D., *Desulfitobacterium*; Dv., *Desulfovibrio*; G., *Geobacter*; Gl., *Gluconacetobacter*; H., *Herbaspirillum*; K., *Klebsiella*; M., *Mesorhizobium*; Mb., *Methanobacterium*; Mc., *Methanococcus*; Ms., *Methanosarcina*; R., *Rhizobium*; Rba., *Rhodobacter*; Rsp., *Rhodospirillum*; S., *Sinorhizobium*; T., *Trichodesmium*. Functional nitrogenase groups and the organization of the *nif* operon in each diazotroph are shown on the right; genes other than *nifHDKEN* or *nif1* and *nif2* are marked as open arrows; ~ indicates more than five ORFs.

early origin of these genes, preceding the divergence of the two groups of prokaryotes, and the last common ancestor (LCA) most likely had *nifI* genes in its nitrogenase family. During the divergence between groups I and II, which resulted from the development of oxygenic photosynthesis and the subsequent aerobic/anaerobic segregation of environments, *nifI* genes were lost, although the reason for this remains unclear. Genome analyses showed that *nifI* genes have not been found in aerobic diazotrophs as yet. That the *nifI* genes have an early origin was also assumed to be true in the hypothesis holding that nitrogen fixation first arose in methanogenic archaea (Raymond *et al.*, 2004), since all types of nitrogenases of methanogens in the cluster have *nifI* genes. Surprisingly, the alternative nitrogenases found in aerobic diazotrophs and presumed to have transferred from methanosarcina in both the nitrogen-fixing LCA and methanogen-origin hypothesis (Raymond *et al.*, 2004) did not carry *nifI* genes. This finding again suggests the influence of oxygen on *nifI* evolution.

The *nifI* gene linkage and amino acid sequence similarity suggest that the *nifI* products are most likely involved in switch-off regulation in *Hbt. chlorum* (and other bacteria that carry them), as in *Methanococcus maripaludis*. Actually,

nitrogenase switch-off regulation was reported in *Hbt. chlorum* (Kimble & Madigan, 1992) and *Chlorobium tepidum* (Wahlund & Madigan, 1993), although the mechanism or proteins involved are still unknown. This regulation was also observed in *Clostridium beijerinckii* and it was speculated that products of *nifI* genes play a role in it, since *Clostridium pasteurianum*, which lacks *nifI* genes in the genome, did not show similar regulation of nitrogenase to ammonia (Chen, 2004). Although NifI proteins are highly similar to GlnB and GlnK, they differ significantly in the T-loop region where the interaction with other proteins occurs. The conserved site (within the T-loop) for uridylylation or phosphorylation is absent in NifI (Kessler *et al.*, 2001), and this is also true for the *Hbt. chlorum* and *D. hafniense* NifI. This suggests that *Hbt. chlorum* NifI might have functional similarity to the archaeal NifI, although the detailed mechanism is still unknown.

Conclusions

The *Hbt. chlorum* *nif* gene cluster was concise, and consisted of 11 genes arranged within a 10 kb region (Fig. 1). Actually, among the diazotrophs, the smallest number of genes required for nitrogen fixation has been found in

Fig. 4. Phylogenetic trees for *Nifl* created with the NJ and ML methods. Sequences of *GlnB* and *GlnK*, close relatives of *Nifl*, are used as an outgroup. Corresponding accession numbers are indicated next to the bacterial names. Bootstrap values above 50 % are shown for each node. Sequences obtained in this study and branches supported by bootstrap values of 90 % or more are shown in bold. Sequences other than *Nifl*/*GlnB* are indicated by: *, *GlnK*; a, Anf; v, Vnf. Genus abbreviations not given in the legend to Fig. 3: *B.*, *Bacillus*; *E.*, *Escherichia*; *Str.*, *Streptomyces*.

strictly anaerobic prokaryotes; for example, the *Methanococcus maripaludis nif* cluster contains eight genes (Kessler *et al.*, 1998) and the *nif* cluster of *Clostridium acetobutylicum* consists of nine genes (Chen, 2004) (Fig. 5). The universal

presence of *nifl* genes in strictly anaerobic prokaryotes suggested the essential role of these genes in nitrogen fixation, probably in the regulation of nitrogenase protein as in *Methanococcus maripaludis*. The relatively ‘simple’

Fig. 5. Comparison of the *nif* cluster in representatives of the nitrogenase groups. UAS, Upstream activator sequence; -24 -12, promoter recognized by σ^{54} . Functions of regulatory elements are shown in boxes and predicted functions are shown in dotted boxes. For *Azotobacter vinelandii*, the gene and ORF designations are as presented by Jacobson *et al.* (1989).

nitrogenase system of the strict anaerobes is regulated by NifH, although the mechanism is unknown. On the contrary, the extensive *nif* clusters of aerobic diazotrophs, for example the *nif* cluster from *Azotobacter vinelandii*, lack *nifH* genes but have acquired other regulatory genes (Fig. 5). The highly conserved nitrogen-fixation cluster in *Hbt. chlorum* was found to be acquired through vertical transfer from the LCA or the methanogen, and its phylogenetic position as an intermediate between group I, consisting of aerobic diazotrophs, and group II, consisting of strictly anaerobic prokaryotes, may reflect an evolutionary stage of a divergence of the two nitrogenase groups.

ACKNOWLEDGEMENTS

This study was supported in part by a grant from IFO (Institute for Fermentation, Osaka, Japan) to H.K. This paper represents a portion of the dissertation submitted by J. E.-A. to Osaka University in partial fulfilment of the requirements for a PhD degree.

REFERENCES

- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–3402.
- Arcondeguy, T., Jack, R. & Merrick, M. (2001). P_{II} signal transduction proteins, pivotal players in microbial nitrogen control. *Microbiol Mol Biol Rev* **65**, 80–105.
- Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K. (editors) (1995). *Current Protocols in Molecular Biology*, vol. 1. New York: Wiley.
- Beer-Romero, P. & Gest, H. (1987). *Heliobacillus mobilis*, a peritrichously flagellated anoxyphototroph containing bacteriochlorophyll *g*. *FEMS Microbiol Lett* **41**, 109–114.
- Brigle, K. E., Newton, W. E. & Dean, D. R. (1985). Complete nucleotide sequence of the *Azotobacter vinelandii* nitrogenase structural gene cluster. *Gene* **37**, 37–44.
- Brockmann, H., Jr & Lipinski, A. (1983). Bacteriochlorophyll *g*: a new bacteriochlorophyll from *Heliobacterium chlorum*. *Arch Microbiol* **136**, 17–19.
- Bryantseva, I. A., Gorlenko, V. M., Kompantseva, E. I., Achenbach, L. A. & Madigan, M. T. (1999). *Heliorestis daurensis*, gen. nov. sp. nov. an alkaliphilic rod-to-coiled-shaped phototrophic heliobacterium from a Siberian soda lake. *Arch Microbiol* **172**, 167–174.
- Bryantseva, I. A., Gorlenko, V. M., Kompantseva, E. I., Tourova, T. P., Kuznetsov, B. B. & Osipov, G. A. (2000). Alkaliphilic heliobacterium *Heliorestis baculata* sp. nov. and emended description of the genus *Heliorestis*. *Arch Microbiol* **174**, 283–291.
- Cantera, J. J. L., Kawasaki, H. & Seki, T. (2004). The nitrogen-fixing gene (*nifH*) of *Rhodopseudomonas palustris*: a case of lateral gene transfer? *Microbiology* **150**, 2237–2246.
- Chen, J.-S. (2004). Nitrogen fixation in the clostridia. In *Genetics and Regulation of Nitrogen Fixation in Free-Living Bacteria*, pp. 53–64. Edited by W. Klipp, B. Masepohl, J. R. Gallon & W. E. Newton. Dordrecht & London: Kluwer Academic.
- Chen, J.-S., Toth, J. & Kasap, M. (2001). Nitrogen-fixation genes and nitrogenase activity in *Clostridium acetobutylicum* and *Clostridium beijerinckii*. *J Ind Microbiol Biotechnol* **27**, 281–286.
- Chien, Y. T. & 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.
- Enkh-Amgalan, J., Kawasaki, H. & Seki, T. (2005). NifH and NifD sequences of heliobacteria: a new lineage in the nitrogenase phylogeny. *FEMS Microbiol Lett* **243**, 73–79.
- Felsenstein, J. (2004). PHILIP – Phylogeny Inference Package, version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle, USA.
- Gest, H. & Favinger, J. L. (1983). *Heliobacterium chlorum*, an anoxygenic brownish-green photosynthetic bacterium containing a “new” form of bacteriochlorophyll. *Arch Microbiol* **136**, 11–16.
- Harriott, O. T., Hosted, T. J. & Benson, D. R. (1995). Sequences of *nifX*, *nifW*, *nifZ*, *nifB* and two ORF in the *Frankia* nitrogen fixation gene cluster. *Gene* **161**, 63–67.
- Henson, B. J., Watson, L. E. & Barnum, S. R. (2004). The evolutionary history of nitrogen fixation, as assessed by *nifD*. *J Mol Evol* **58**, 390–399.
- Jacobson, M. R., Brigle, K. E., Bennett, L., Setterquist, R. A., Wilson, R. A., Cash, V. L., Beynon, J., Newton, W. E. & Dean, D. R. (1989). Physical and genetic map of the major *nif* gene cluster from *Azotobacter vinelandii*. *J Bacteriol* **171**, 1017–1027.
- Jeanmougin, F., Thompson, J. D., Gouy, M., Higgins, D. G. & Gibson, T. J. (1998). Multiple sequence alignment with CLUSTAL X. *Trends Biochem Sci* **23**, 403–405.
- Kessler, P. S. & Leigh, J. A. (1999). Genetics in nitrogen regulation in *Methanococcus maripaludis*. *Genetics* **152**, 1343–1351.
- Kessler, P. S., Blank, C. & Leigh, J. A. (1998). The *nif* gene operon of the methanogenic archaeon *Methanococcus maripaludis*. *J Bacteriol* **180**, 1504–1511.
- Kessler, P. S., Daniel, C. & Leigh, J. A. (2001). Ammonia switch-off of nitrogen fixation in the methanogenic archaeon *Methanococcus maripaludis*: mechanistic features and requirement for the novel GlnB homologues, Nif_I and Nif_L. *J Bacteriol* **183**, 882–889.
- Kimble, L. K. & Madigan, M. T. (1992). Nitrogen fixation and the nitrogen metabolism in heliobacteria. *Arch Microbiol* **158**, 155–161.
- Kimble, L. K., Mandelco, L., Woese, C. R. & Madigan, M. T. (1995). *Heliobacterium modesticaldum*, sp. nov., a thermophilic heliobacterium of hot springs and volcanic soils. *Arch Microbiol* **163**, 259–267.
- Kimble-Long, L. K. & Madigan, M. T. (2001). Molecular evidence that the capacity for endospore formation is universal among phototrophic heliobacteria. *FEMS Microbiol Lett* **199**, 191–195.
- Lee, S., Reth, A., Meletzus, D., Sevilla, M. & Kennedy, C. (2000). Characterization of a major cluster of *nif*, *fix*, and associated genes in a sugarcane endophyte, *Acetobacter diazotrophicus*. *J Bacteriol* **182**, 7088–7091.
- Madigan, M. T. (1992). The family Heliobacteriaceae. In *The Prokaryotes*, 2nd edn, pp. 1981–1992. Edited by A. Balows, H. G. Trüper, M. Dworkin, W. Harder & K.-H. Schleifer. Berlin, Heidelberg & New York: Springer.
- Madigan, M. T. (2001). Family VI. “Heliobacteriaceae” Beer-Romero and Gest 1987, 113. In *Bergey’s Manual of Systematic Bacteriology*, 2nd edn, vol. 1, pp. 625–626. Edited by D. R. Boone, R. W. Castenholz & G. M. Garrity. New York: Springer.
- Minerdi, D., Fani, R., Gallo, R., Boarino, A. & Bonfante, P. (2001). Nitrogen fixation genes in an endosymbiotic *Burkholderia* strain. *Appl Environ Microbiol* **67**, 725–732.
- Niggemeyer, A., Spring, S., Stackebrandt, E. & Rosenzweig, R. F. (2001). Isolation and characterization of a novel As(V)-reducing bacterium: implications for arsenic mobilization and the genus *Desulfitobacterium*. *Appl Environ Microbiol* **67**, 5568–5580.

- Ninfa, A. J. & Atkinson, M. R. (2000). PII signal transduction proteins. *Trends Microbiol* **8**, 172–179.
- Noda, S., Ohkuma, M., Usami, R., Horikoshi, K. & 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.
- Norel, F. & Elmerich, C. (1987). Nucleotide sequence and functional analysis of the two *nifH* copies of *Rhizobium* ORS571. *J Gen Microbiol* **133**, 1563–1576.
- Oppenheim, D. S. & Yanofsky, C. (1980). Translational coupling during expression of the tryptophan operon of *Escherichia coli*. *Genetics* **95**, 785–795.
- Ormerod, J. G., Kimble, L. K., Nesbakken, T., Torgersen, Y. A., Woese, C. R. & Madigan, M. T. (1996). *Heliophilum fasciatum* gen. nov. sp. nov. and *Heliobacterium gestii* sp. nov.: endospore-forming heliobacteria from rice field soils. *Arch Microbiol* **165**, 226–234.
- Page, R. D. M. (1996). TreeView: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci* **12**, 357–358.
- Perrière, G. & Gouy, M. (1996). WWW-query: an on-line retrieval system for biological sequence banks. *Biochimie* **78**, 364–369.
- Raymond, J., Siefert, J. L., Staples, C. R. & Blankenship, R. E. (2004). The natural history of nitrogen fixation. *Mol Biol Evol* **21**, 541–554.
- Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**, 406–425.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* **74**, 5463–5467.
- Sibold, L., Henriquet, M., Possot, O. & Aubert, J. P. (1991). Nucleotide sequence of *nifH* regions from *Methanobacterium ivanovii* and *Methanosarcina barkeri* 227 and characterization of *glnB*-like genes. *Res Microbiol* **142**, 5–12.
- Stevenson, A. K., Kimble, L. K. & Madigan, M. T. (1997). Characterization of new heliobacteria and their habitats. *Photosynth Res* **53**, 1–12.
- Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997). The CLUSTAL_X Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**, 4876–4882.
- Wahlund, T. M. & Madigan, M. T. (1993). Nitrogen fixation by the thermophilic green sulfur bacterium *Chlorobium tepidum*. *J Bacteriol* **175**, 474–478.
- Young, J. P. W. (1992). Phylogenetic classification of nitrogen-fixing organisms. In *Biological Nitrogen Fixation*, pp. 43–86. Edited by G. Stacey, R. H. Burris & H. J. Evans. New York: Chapman & Hall.
- Zehr, J. P., Jenkins, B. D., Short, S. M. & Steward, G. F. (2003). Nitrogenase gene diversity and microbial community structure: a cross-system comparison. *Environ Microbiol* **5**, 539–554.