J. Genet. Vol. 69, No. 2, August 1990, pp. 67–78. © Printed in India.

Cluster analysis of genes for nitrogen fixation from several diazotrophs

MANJULA MATHUR and RAKESH TULI*

Molecular Biology and Agriculture Division, Bhabha Atomic Research Centre, Trombay, Bombay 400 085, India

MS received 27 January 1990

Abstract. Hierarchical clustering and similarity coefficients of pairwise alignments of the published nucleotide sequences of 27 *nif*H genes suggest that *nif* genes are as ancient as the archaebacteria and clostridia. The positions of *nif*H1 of *Methanococcus thermolithotrophicus*, *nif*H3 of *Clostridium pasteurianum*, *nif*H3 of *Azotobacter vinelandii* and *nif*H of *Frankia* suggest that a variety of lateral transfers may have occurred during evolution of *nif*H gene. The genes for type 3 nitrogenase of *A. vinelandii* may have diverged early from methanogens and clostridia. A high similarity coefficient with the derived amino acid sequence of type 3 nitrogenase suggests the presence of a functionally similar enzyme in *C. pasteurianum*. The type 2 nitrogenase gene *nif*H2 of azotobacters seems to have originated recently from the gene *nif*H1 for conventional type 1 nitrogenase. Rhizobial *nif*H genes comprise two closely related but discrete clusters that are in consonance with the plasmid or chromosomal location of *nif* genes. The chromosomal and plasmid located *nif*H of rhizobia seem to have evolved independently but contemporaneously.

Keywords. Diazotrophs; *nif* genes; phylogeny; sequence analysis; dendrogram; evolution; nitrogenase.

1. Introduction

The ability to reduce atmospheric dinitrogen into ammonia is found in a number of phylogenetically diverse prokaryotes including archaebacteria. Biological nitrogen fixation is catalysed by a complex enzyme nitrogenase, essentially composed of two components. Component I is an $\alpha_2\beta_2$ tetramer while component II is a dimer comprising two identical subunits coded by *nif*H (Eady and Smith 1979). Subunits α and β of component I are coded by *nif*D and *nif*K respectively. *nif*H is highly conserved among all diazotrophs (Ruvkun and Ausubel 1980). Nucleotide sequences of regions homologous to *nif*H have been reported in literature from at least 18 diazotrophs. Multiple *nif*H have been reported in many of those including *Clostridium pasteurianum* (Wang *et al.* 1988), *Rhodobacter capsulatus* (Scolnik and Haselkorn 1984), *Rhizobium phaseoli* (Quinto *et al.* 1985), *Rhizobium sp.* strain ORS571 (Norel and Elmerich 1987), *Azotobacter vinelandii* (Jacobson *et al.* 1986), *Azotobacter chroococcum* (Robson *et al.* 1986) and *Methanococcus thermolithotrophicus* (Souillard and Sibold 1989). Here we report some interesting observations made by cluster analysis of nucleotide sequences and the derived amino acid sequences of *nif* genes. The preliminary findings were reported by us recently (Mathur and Tuli 1990). While the manuscript was in preparation we noticed that some of our interpretations are similar to those published recently by Normand and Bousquet (1989). However, these authors have based their analysis only on derived amino acid sequences of *nif*H genes.

*For correspondence.

2. Methods

Sources of *nif*H, *nif*D and *nif*K sequences used in the analysis are listed in table 1 along with the name symbols used in the text. The multiple *nif*H genes that have

Table 1. Sources of nucleotide sequences.

Name of the organism	Symbol	Reference
For nifH genes:		
Anabaena sp. strain 7120	AnH	Mevarech *et al.* (1980)
Azorhizobium sp. strain ORS571	AsH1	Norel and Elmerich (1987)
Azorhizobium sp. strain ORS571	AsH2	Norel and Elmerich (1987)
Azotobacter chroococcum	AcH2	Robson *et al.* (1986)
Azotobacter vinelandii	AvH1	Brigle *et al.* (1985)
Azotobacter vinelandii	AvH2	Raina *et al.* (1988)
Azotobacter vinelandii	AvH3	Joerger *et al.* (1989)
Bradyrhizobium japonicum	BjH	Adams and Chelm (1984)
Bradyrhizobium sp. strain ANU289	BsH	Scott *et al.* (1983a)
Clostridium pasteurianum	CpH1	Wang *et al.* (1988)
Clostridium pasteurianum	CpH2	Wang *et al.* (1988)
Clostridium pasteurianum	CpH3	Wang *et al.* (1988)
Clostridium pasteurianum	CpH4	Wang *et al.* (1988)
Clostridium pasteurianum	CpH5	Wang *et al.* (1988)
Clostridium pasteurianum	CpH6	Wang *et al.* (1988)
Frankia sp. strain ArI3	FrH	Normand *et al.* (1988)
Klebsiella pneumoniae	KpH	Sundaresan and Ausubel (1981)
Methanobacterium ivanovii	MiH	Souillard *et al.* (1988)
Methanococcus thermolithotrophicus	MtH1	Souillard and Sibold (1989)
Methanococcus thermolithotrophicus	MtH2	Souillard *et al.* (1988)
Methanococcus voltae	MvH	Souillard and Sibold (1986)
Rhizobium meliloti strain 41	RmH	Toerock and Kondorosi (1981)
Rhizobium phaseoli	RpH	Quinto *et al.* (1985)
Rhizobium sp. strain ANU240	RsH	Jones *et al.* (1989)
Rhizobium trifolii strain 329	RtH	Scott *et al.* (1983b)
Rhodobacter capsulatus	RcH	Jones and Haselkorn (1988)
Thiobacillus ferrooxidans	TfH	Pretorius *et al.* (1987)
For nifD genes:		
Anabaena sp. strain 7120	AnD	Lammers and Haselkorn (1983)
Azotobacter vinelandii	AvD1	Brigle *et al.* (1985)
Azotobacter vinelandii	AvD3	Joerger *et al.* (1989)
Bradyrhizobium cowpea IRc78	BcD	Yunn and Szalay (1984)
Bradyrhizobium japonicum	BjD	Kaluza and Hennecke (1984)
Bradyrhizobium sp. strain ANU289	BsD	Weinmann *et al.* (1984)
Clostridium pasteurianum	CpD	Wand *et al.* (1987)
Klebsiella pneumoniae	KpD	Ioannidis and Buck (1987)
Methanococcus thermolithotrophicus	MtD1	Souillard and Sibold (1989)
Rhodobacter capsulatus	RcD	Schumann *et al.* (1986)
For nifK genes:		
Anabaena sp. strain 7120	AnK	Mazur and Chui (1982)
Azotobacter vinelandii	AvK1	Brigle *et al.* (1985)
Azotobacter vinelandii	AvK3	Joerger *et al.* (1989)
Bradyrhizobium sp. strain ANU289	BsK	Weinmann *et al.* (1984)
Klebsiella pneumoniae	KpK	Holland *et al.* (1987)

been sequenced from different organisms are given numbers from *nif*H1 to *nif*H6. The *nif*H3 of *A. vinelandii* is the gene for component II of a recently discovered nitrogenase that contains iron but no molybdenum or vanadium (called type 3 nitrogenase). The protein corresponding to *nif*H2 contains Fe and V but no Mo (type 2 nitrogenase) and *nif*H1 makes component II of conventional nitrogenase that contains Fe and Mo (type 1 nitrogenase). In other cases, the functional significance of multiple *nif* genes is not known and orders without any prior assumptions.

Hierarchical clustering of the genes and their derived polypeptide sequences was done by multiple alignment program described by Corpet (1988) using the unite. dat symbol comparison table and gap penalty 7. This program takes into account the closer relationships that can exist among subsets of sequences.

For individual quantitative comparisons of nucleotide as well as peptide sequences, pairwise alignments were done using the program of Wilbur and Lipman (1983). Peptide sequences were aligned with tuple size 1 and gap penalty 1. For aligning nucleotide sequences tuple size was set to 2 in order to take care of variability at the 3rd position of codons and gap penalty was increased to 7 to minimise out-of-frame matching of nucleotides. Similarity coefficient S_{AB} between two sequences, say A and B, was calculated according to Fox *et al.* (1977),

$$S_{AB} = \frac{2 \times \text{no. of matching residues in A and B} \times 100}{(\text{total no. of residues in A}) + (\text{total no. of residues in B})}.$$

3. Results and discussion

3.1 *Type 3 nitrogenase genes of A. vinelandii cluster with primitive nif genes*

Hierarchical clustering of *nif*H nucleotide sequences obtained after 2 iterations of multiple alignment is shown in figure 1. *M. thermolithotrophicus* carries two *nif*H genes, of which only *nif*H1 is known to be functional (Souillard and Sibold 1989). The other *nif*H of *M. thermolithotrophicus*, named MtH2 clusters with the other two methanogens namely *Methanococcus voltae* and *Methanobacterium ivanovii* which are actually quite distant from each other. Methanogens are anaerobic archaebacteria which are generally agreed to be early habitants of earth (Fox *et al.* 1977). The relatively high degree of dissimilarity among *nif*H genes of methanogens is also reflected in low similarity coefficients (table 2) among these. Clustering of the *nif* genes of the two methanococci and their divergence from *Methanobacterium* is in agreement with the phylogeny of methanogens proposed on the basis of 16S and 5S rRNA comparisons (Woese 1987). From methanogens the first bifurcation is towards *Clostridium pasteurianum*, an anaerobic gram-positive eubacterium on one side and aerobic diazotrophs on the other. *C. pasteurianum* has been reported to have several *nif*H genes, of which six have been sequenced (Wang *et al.* 1988). CpH1 is reported to be transcriptionally most active. *nif*H3 is very different from all other *nif*H of *C. pasteurianum* in that it clusters with *nif*H1 of *M. thermolithotrophicus* rather than with the other five *C. pasteurianum nif*H genes. Wang *et al.* (1988) reported cluster analysis for *nif*H genes of *C. pasteurianum* and concluded that CpH5, CpH2, CpH6 and CpH4 are closely related to CpH1. Our detailed analysis shows that CpH3 is further separated from other *nif*H genes of *C. pasteurianum* and

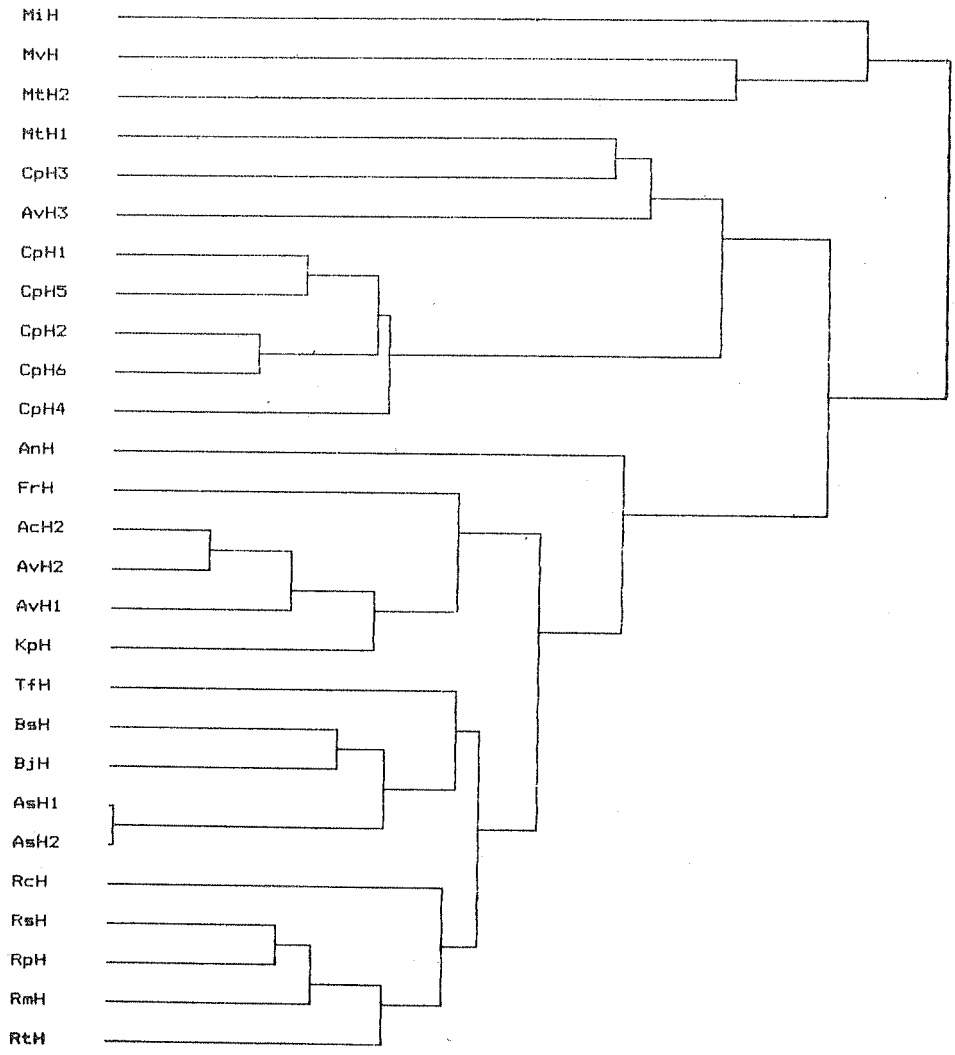

Figure 1. Hierarchical clustering of nucleotide sequences of *nif*H.

may have originated independently along with the *nif*H1 of *M. thermolithotrophicus.* AvH3 is known to form component II of a recently discovered functional nitrogenase in *A. vinelandii* under conditions of molybdenum and vanadium deficiency, a function which has not yet been discovered for *C. pasteurianum* or *M. thermolithotrophicus.* However, reduction of acetylene to ethane has been reported in *C. pasteurianum* (Dilworth *et al.* 1987), suggesting the possible presence of alternative nitrogenase. Quite significantly the *nif*H3 of *A. vinelandii* clusters with *nif*H3 of *C. pasteurianum* rather than with *nif*H1 and *nif*H2 of azotobacters. Clustering suggests that the evolutionary position of *nif*H genes for type 1 or Mo-nitrogenase (AvH1) and type 2 or V-nitrogenase (AvH2 and AcH2) of azotobacters is quite different from that for type 3 or Fe-nitrogenase (AvH3). The V-nitrogenase appears to have originated much later and independently of the Fe-nitrogenase. A

Table 2. Similarity coefficient (S_{AB}) matrix calculated by pairwise alignments of nucleotide sequences (above the stars '*'), and derived amino acid sequences (below the stars '*').
For *nif*H genes:

	Mt1	Cp3	Av3	Cp1	An	Fr	Kp	Ac2	Av1	Tf	Bs	Bj	As	Rc	Rt	Rp	Rm	Rs	Mt2	Mv	Mi
Mt1	*	68	60	65	59	54	57	50	57	49	55	46	54	48	56	56	56	55	52	58	57
Cp3	70	*	61	68	58	52	56	47	57	46	54	49	52	43	54	54	55	53	50	57	58
Av3	71	82	*	48	60	58	61	68	62	61	57	61	59	66	55	55	58	61	29	41	46
Cp1	64	65	62	*	58	52	56	47	58	48	54	49	52	48	55	54	54	53	57	58	59
An	59	58	58	61	*	68	66	67	68	58	61	60	63	59	63	64	64	61	50	51	47
Fr	59	61	57	62	77	*	72	75	78	65	70	65	74	68	68	71	70	72	38	40	40
Kp	62	63	61	65	71	75	*	80	80	69	71	66	73	69	68	70	67	70	48	51	49
Ac2	61	62	62	66	72	76	85	*	88	71	68	70	72	75	67	70	68	70	38	41	34
Av1	62	62	61	67	71	76	88	90	*	69	72	69	73	72	69	72	70	71	48	51	50
Tf	59	57	58	61	72	72	76	72	74	*	72	74	75	69	66	67	65	69	38	40	40
Bs	58	58	57	62	71	73	76	73	75	86	*	88	82	69	71	74	73	76	46	48	48
Bj	57	58	57	61	72	73	76	74	75	86	97	*	78	71	66	68	67	69	39	43	40
As	58	58	57	61	72	72	75	71	73	84	93	92	*	73	72	75	73	77	33	48	47
Rc	59	59	60	65	70	71	73	73	74	78	78	77	77	*	71	73	71	75	29	42	39
Rt	58	55	58	62	71	72	69	71	70	78	77	78	77	80	*	71	80	81	47	49	43
Rp	59	56	58	62	72	72	70	72	69	79	80	80	78	81	90	*	84	88	48	49	44
Rm	59	57	57	62	71	71	69	71	69	79	78	79	78	80	91	94	*	85	47	48	46
Rs	59	58	57	62	70	70	69	71	69	78	79	78	77	81	89	93	92	*	33	47	46
Mt2	47	48	47	51	46	47	49	47	48	46	46	46	45	47	47	46	48	46	*	58	53
Mv	53	52	52	52	50	49	50	51	50	47	48	49	48	47	50	50	49	48	54	*	58
Mi	52	51	52	54	47	49	49	51	49	46	48	48	47	48	50	48	47	48	50	53	*

For *nif*D genes:

	Mt2	Av3	Cp1	An	Kp	Av1	Bs	Bj	Bc	Rc
Mt2	*	37	33	42	42	41	42	41	43	24
Av3	39	*	28	35	40	41	37	37	38	31
Cp1	41	35	*	31	31	26	28	28	27	37
An	39	31	38	*	45	50	61	60	61	39
Kp	42	35	43	63	*	72	59	57	58	58
Av1	41	33	41	63	72	*	69	66	68	58
Bs	42	35	39	66	69	72	*	87	90	50
Bj	42	36	40	65	69	70	92	*	88	61
Bc	42	36	38	65	69	71	93	91	*	62
Rc	37	34	39	63	64	64	69	67	68	*

For *nif*K genes:

	Av3	An	Kp	Av1	Bs
Av3	*	27	31	34	27
An	35	*	53	47	50
Kp	32	53	*	58	47
Av1	34	55	66	*	48
Bs	34	59	51	54	*

The symbols in the first row and first column of the matrix are the name symbols for the dizotrophs – identical to the symbols in table 1 but without the third letter. AvH2 is not included in the matrix. Its amino acid sequence is 96% similar to amino acid sequence of AcH2 while their nucleotide sequences are 90% similar. Its similarity coefficients with all other sequences are similar to those for AcH2.

similar observation has been made by Normand and Bousquet (1989) also. The derived amino acid sequences of the products of the two *nif*H3 genes show a much higher degree of similarity (S_{AB} 82%, table 2, figure 2) suggesting the possible existence of a functionally conserved protein in *C. pasteurianum*. The other two genes *nif*D3 (figure 3) and *nif*K3 (figure 4) for type 3 nitrogenase of *A. vinelandii* also

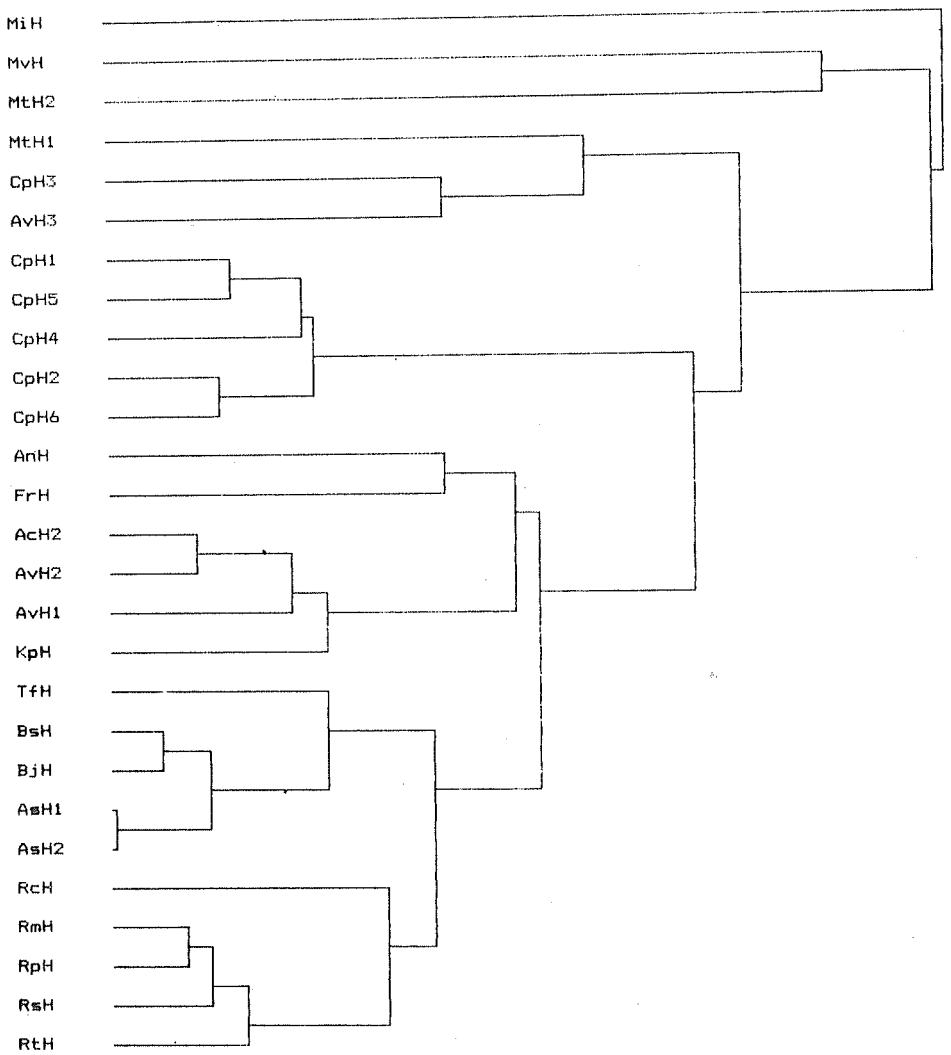

Figure 2. Hierarchical clustering of derived amino acid sequences of *nif*H.

cluster separately from the corresponding genes for type 1 nitrogenase substantiating independent and early evolution of the type 3 nitrogenase genes.

3.2 *Cyanobacterial nif genes emerge at the interphase of anaerobes and aerobes*

After the first bifurcation from methanogens, *Anabaena* is the next to evolve. In the cluster diagram for *nif*H (figure 1) *Anabaena* 7120 occupies the position after methanogens and *C. pasteurianum*, agreeing with the suggestion that cyanobacteria evolved at the interphase of anaerobes and aerobes. In this respect, the position of *Anabaena nif*D is similar to that of *nif*H in the hierarchical clustering of *nif*D sequences of 10 diazotrophs (figure 3). The *Anabaena nif*H and *nif*D remain quite

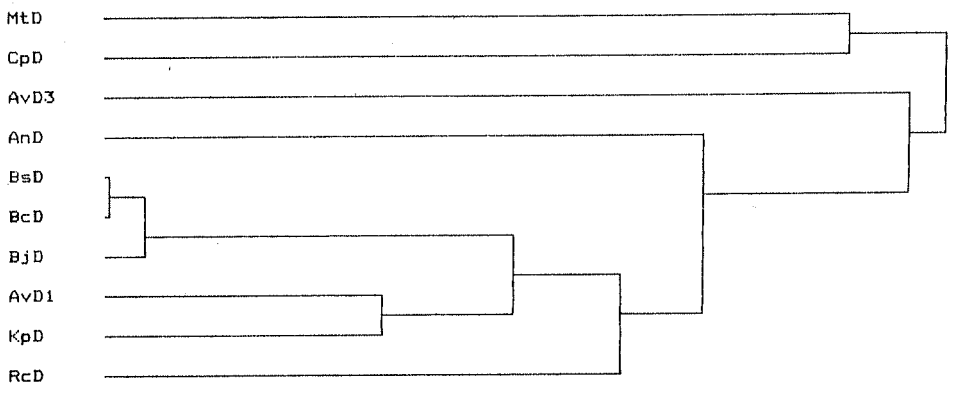

Figure 3. Hierarchical clustering of nucleotide sequences of *nif*D.

distant from other bacteria suggesting that the diazotrophic cyanobacteria may not have exchanged *nif* genes with other bacteria. The data available for *nif*K is rather limited as only five sequences have been published (figure 4). The hierarchical position of *Anabaena nif*K is quite different suggesting that *nif*K may have evolved differently from *nif*H and *nif*D.

3.3 *Clustering of nif genes among aerobes*

The *nif*H genes of aerobic diazotrophs are bifurcated into two main groups in hierarchical clustering. The first group includes gram-positive actinomycetes *Frankia* along with azotobacters and *K. pneumoniae*. Since *Frankia* is gram-positive like clostridia, the separation of its *nif*H from *C. pasteurianum* was considered surprising in the recent cluster analysis done by Normand and Bousquet (1989) on the basis of derived amino acid sequences. However, the cluster diagram drawn by us on the basis of nucleotide sequences (figure 1) shows that the separation of these two sequences is in agreement with the differences between the genomic G+C content of the two organisms, the average G+C of *Frankia* being 66–68% and that of *C. pasteurianum* being 26–28%. However, in phylogenies based on both 16S rRNA (Woese 1987) and 5S rRNA (Dams *et al.* 1987), actinomyces branch away from cyanobacteria. This point has again been discussed later. Our analysis includes the *nif*H sequence from the *Frankia* strain ArI3 which carries it on the chromosome. While the manuscript was in preparation, another chromosome-borne *nif*H sequence from the *Frankia* strain HRN18a was reported by Normand and

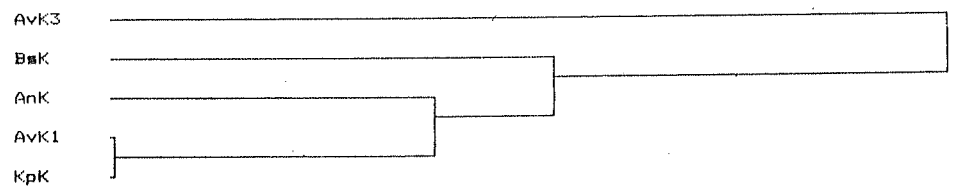

Figure 4. Hierarchical clustering of nucleotide sequences of *nif*K.

Bousquet (1989). The nucleotide sequence is 93% similar to *nif*H of *Frankia* strain ArI3 and would cluster with it in our analysis also. Some strains of *Frankia* have been reported to carry plasmid-borne *nif* genes (Normand *et al.* 1988). However, none of such *nif* genes have been sequenced. It would be interesting to examine if the extrachromosal *nif* genes of *Frankia* show a phylogenetic relationship different from its chromosomal *nif* genes.

The group of aerobic eubacteria is further bifurcated in figure 1. In one of the branches some rhizobia bifurcate alongwith the beta purple bacterium *Thiobacillus ferrooxidans*. In the other branch, several other rhizobial *nif*H bifurcate alongwith the alpha purple bacterium *Rhodobacter capsulatus*. The clustering of nucleotide sequences (figure 1) is quite similar to that of the derived amino acid sequences (figure 2) in this respect. In the classification by Woese (1987) done on the basis of 16S rRNA comparisons, rhizobia and *Rhodobacter capsulatus* are grouped as alpha purple bacteria. Azotobacters and enterics are gamma purple bacteria. In view of evolution of bacterial respiration, Broda (1971) suggested that strict aerobes (like azotobacters) could be derived from the facultatively aerobic coliforms, (such as *Klebsiella pneumoniae*). Purple bacteria are the phylum whose ancestral phenotype is photosynthetic (purple). Photosynthetic capacity has been lost many times in this phylum, resulting in various nonphotosynthetic sublines. Thus clustering of gram-negative bacteria on the basis of *nif* genes is similar to that on the basis of 16S rRNA.

3.4 *Rhizobial nif genes cluster in accordance with gene location*

The clustering of *nif*H genes of rhizobia into two groups is not in concordance with either host range or growth characteristics. However, the clustering is in complete agreement with the location of *nif* genes. For instance, the stem nodule forming *Azorhizobium* has chromosomal *nif* genes but shows growth characteristics similar to fast growing rhizobia (Donald *et al.* 1986). However, it forms a cluster with bradyrhizobia (figure 1) which have chromosomal *nif* genes and grow slowly. Similarly *Rhizobium sp.* strain ANU240 has a very broad host range (Jones *et al.* 1989) unlike the other rhizobia with which it clusters. However, all the rhizobia in this cluster have plasmid borne *nif* genes. Cluster analysis based on peptide sequence similarity (figure 2) also shows similar results. The *nif*D gene has been sequenced from only three rhizobia. In all the three cases, *nif* genes are located on chromosome. These show very high sequence similarity (table 2) and cluster together (figure 3). None of the plasmid borne *nif*D or *nif*K genes has been sequenced. Based on several characters, bradyrhizobia have been recently proposed to form a genus separate from the fast growing rhizobia. The genome size of fast-growing rhizobia is almost half that of bradyrhizobia (Chakrabarti *et al.* 1984). Deletion of genomic DNA is suggested to have occurred from bradyrhizobia resulting in the removal of information that was no longer required in the new environment to give rise to fast-growing rhizobia. Immunological comparison of nitrogenase (Bisseling *et al.* 1982) suggested that fast-growing and slow-growing rhizobia are as far apart from each other as they are from *K. pneumoniae*. Unlike in the case of rhizobia and most of the other diazotrophs, *nif*H and *nif*DK are not contiguous on the genome of bradyrhizobia. However, *nif*H cluster analysis does

not agree with the erstwhile used parameters for classifying rhizobia. The cluster analysis suggests that there has not been a recent exchange of *nif* genes between rhizobia carrying plasmid-borne *nif* genes and those with chromosomally located *nif* genes. Our analysis suggests that the plasmid-borne and the chromosomal *nif*H in rhizobia evolved contemporaneously from a common progenitor. Sequencing of more rhizobial *nif* genes in future would be helpful in establishing these points.

The positions of *Anabaena* and *Bradyrhizobium* (Parasponia) *sp ANU289 nif*K genes (figure 4) are different from those in cluster diagrams for *nif*H and *nif*D suggesting that *nif*K may have evolved differently from *nif*H and *nif*D. In general S_{AB} values for *nif*D sequences are less than those for *nif*H sequences and S_{AB} values for *nif*K sequences are the least (table 2). This puts the *nif*H product at a key position in nitrogenase function with *nif*D also at a functionally important position while variability in *nif*K may have been permitted to take care of different physiological conditions. Thus differences in mutational pressures may have led to the differences in phylogenetic relationships.

3.5 *Nucleotide sequence vs amino acid sequence based cluster analysis for translatable genes*

Normand and Bousquet (1989) did not use nucleotide sequences for distance matrix analysis due to the high G+C skew at the third codon positions of *nif*H genes. A high G+C skew at third codon positions indicates the existence of high GC/AT mutation pressure (Osawa *et al.* 1987) on *nif*H. We have earlier suggested (Mathur and Tuli 1989) that the widely different G+C content of *nif* genes (34% to 65%) reflects adaptability in terms of codon usage, that may have occurred gradually during evolution, to achieve a functionally appropriate level of expression of the genes in the respective cellular milieu. The G+C content at the third codon position has been reported to have a strong bearing on codon usage in different organisms (Osawa *et al.* 1987). Our detailed analysis of codon usage (to be published elsewhere) shows that the G+C content of type 3 nitrogenase genes of *A. vinelandii* is not different from that of type 1 and type 2 genes. Inspite of similar GC pressure, AvH3 and AvD3 cluster with *C. pasteurianum* and methanogens rather than with the type 1 and type 2 nitrogenase genes of azotobacters. This strongly suggests that inspite of similar GC pressures on all the *nif* genes of azotobacters, nucleotide sequence comparison shows that distinct phylogenetic identities are retained. In other words, the lineage of *A. vinelandii nif* genes in nucleotide sequence comparison is not effected by the skewed GC content at third codon positions. As seen in figures 1 and 2, phylogenetic position of the various *nif* genes of azotobacters, clostridia and methanogens remain similar in both nucleotide and amino acid sequence comparisons. It is possible that the phylogenetic relationships derived by comparison of nucleotide sequences of translatable genes would show differences as compared to those derived on the basis of non translatable genes, like those for rRNAs. This hypothesis remains to be tested, as mutational pressures on the two classes of genes could be of different nature. Nevertheless, Osawa *et al.* (1987) found that there was a general positive correlation between total G+C content of genomic DNA of various bacteria and that of rRNA genes, tRNA genes, translatable sequences and spacers though the extent of correlation was different.

Grouping of *nif*H by nucleotide sequence comparison (figure 1) agrees well with the cluster diagram drawn on the basis of amino acid sequence comparison (figure 2). However, a noticeable difference in branching order is seen in the position of *Anabaena*. The cyanobacterium branches out distinctly and the present day aerobes emerge later in the nucleotide sequence based cluster analysis. *Anabaena* occupies a position clearly at the interphase of anaerobes and aerobes. However, in amino acid sequence based analysis, *Anabaena* and *Frankia* branch out together and the event is contemporaneous with the branching of several aerobes. This branching order is not in concordance with the phylogenies proposed on the basis of other evidences. Cyanobacteria are taken by many to be more primitive than aerobes. Aerobic respiratory metabolism evolved in bacteria only after the atmosphere became oxygenic following the advent of cyanobacteria which had the ability of oxygenic photosynthesis. On the basis of these arguments, we feel that despite the G+C skew, nucleotide sequence-based cluster analysis of *nif*H may be more representative of phylogenetic relationships than the amino acid-based cluster analysis.

References

Adams T. H. and Chelm B. K. 1984 The *nif*H and *nif*DK promoter regions for *Rhizobium japonicum* share structural homologies with each other and with nitrogen-regulated promoters from other organisms. *J. Mol. Appl. Genet.* 2: 392–405

Arnold W., Rump A., Klipp W., Priefer U. B. and Puhler A. 1988 Nucleotide sequence of a 24,206-base-pair DNA fragment carrying the entire nitrogen fixation gene cluster of *Klebsiella pneumoniae*. *J. Mol. Biol.* 203: 715–738

Bisseling T., van den Bos R. C., Moen L., Hontelez J. G. J. and Kammen van A. 1982 An immunological comparison of nitrogenase proteins of fast and slow growing rhizobia. *FEBS Lett.* 145: 45–48

Brigle K. E., Newton W. E. and Dean D. R. 1985 Complete nucleotide sequence of the *Azotobacter vinelandii* nitrogenase structural gene cluster. *Gene* 37: 37–44

Broda E. 1971 Origins of bacterial respiration. In *Chemical evolution and the origins of life* (eds) R. Buvet and C. Ponnamperumal (Amsterdam: North-Holland) pp. 446–452

Chakrabarti S. K., Mishra A. K. and Chakrabarti P. K. 1984 Genome size variation of rhizobia. *Experientia.* 40: 1290

Corpet F 1988 Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.* 16: 10881–10890

Dams E., Yamada T., Baere R. D., Huysmans E., Vandenberghe A. and Wachter R. D. 1987 Structure of 5S rRNA in Actinomycetes and relatives and evolution of eubacteria. *J. Mol. Evol.* 25: 255–260

Dilworth M. J., Eady R. R., Robson R. L. and Miller R. W. 1987 Ethane formation from acetylene as a potential test for vanadium nitrogenase in vivo. *Nature* (*London*) 327: 167–168

Donald R. G. K., Nees D. W., Raymond C. K., Loroch A. I. and Ludwig R. A. 1986 Characterisation of three genomic loci encoding *Rhizobium Sp.* strain ORS571 N fixation genes. *J. Bacteriol.* 165: 72–81

Eady R. R. and Smith B. E. 1979 Physico-chemical properties of nitrogenase and its components. In *A treatise on dinitrogen fixation, sections I and II* (eds) R. W. F. Hardy, F. Bottomley and R. C. Burns (London: Wiley and Sons) pp. 399–490

Fox G. E., Magrum L. J., Balch W. E., Wolfe R. S. and Woese C. R. 1977 Classification of methanogenic bacteria by 16S ribosomal RNA characterisation. *Proc. Natl. Acad. Sci. USA* 74: 4537–4541

Holland D., Zilberstein A., Zamir A. and Sussman J. L. 1987 A quantitative approach to sequence comparisions of nitrogenase MoFe protein α- and β-subunits including the newly sequenced *nif*K gene from *Klebsiella pneumoniae. Biochem. J.* 247: 277–285

Ioannidis I. and Buck M. 1987 Nucleotide sequence of the *Klebsiella pneumoniae nif*D gene and predicted amino acid sequence of the alpha subunit of the Mo-Fe protein. *Biochem. J.* 247: 287–291

Jacobson M. R., Premkumar R. and Bishop P. E. 1986 Transcriptional regulation of nitrogen fixation by molybdenum in *Azotobacter vinelandii. J. Bacteriol.* 170: 1475–1487

Joerger R. D., Jacobson M. R., Premkumar R., Wolfinger E. D. and Bishop P. E. 1989 Nucleotide sequence and mutational analysis of the structural genes *anf*HDGK for the second alternative nitrogenase from *Azotobacter vinelandii*. *J. Bacteriol.* 171: 1079–1086

Jones J. B., Holton T. A., Morrison C. M., Scott K. F. and Shine J. 1989 Structural and functional analysis of nitrogenase genes from the broad-host-range *Rhizobium* strain ANU240. *Gene* 77: 141–153

Jones R. and Haselkorn R. 1988 DNA sequence of the *Rhodobacter capsulatus nif*H gene. *Nucleic Acids Res.* 17: 8735

Kaluza K. and Hennecke H. 1984 Fine structure analysis of the *nif*DK operon encoding the alpha and beta subunits of dinitrogenase from *Rhizobium japonicum*. *Mol. Gen. Genet.* 196: 35–42

Lammers P. J. and Haselkorn R. 1983 Sequence of the *nif*D gene coding for the alpha subunit of dinitrogenase from the cyanobacterium *Anabaena*. *Proc. Natl. Acad. Sci. USA* 80: 4723–4727

Mathur M. and Tuli R. 1989a *Anabaena* and yeast genes well matched. *Cyanonews* 5: 3

Mathur M. and Tuli R. 1989b General codon usage analysis of functionally related genes from different species. In *Proc. Workshop on Mathematical Models in Biology and Medicine* (Calcutta: Indian Institute of Chemical Biology) p. 81

Mathur M. and Tuli R. 1990a Phylogenetic relationship and codon usage analysis in genes for nitrogen fixation. In *Proc. D.A.E. Symposium on Advances in Molecular Biology*, B.A.R.C., Bombay, pp. 473–474

Mathur M. and Tuli R. 1990b *nif* gene comparison challenges conventional taxonomy. *Cyanonews* 6: 4–5

Mazur B. J. and Chui C. F. 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. and Haselkorn R. 1980 Nucleotide sequence of a cyanobacterial *nif*H gene coding for nitrogenase reductase. *Proc. Natl. Acad. Sci. USA* 77: 6476–6480

Norel F. and Elmerich C. 1987 Nucleotide sequence and functional analysis of two *nif*H copies of *Rhizobium ORS571*. *J. Gen. Microbiol.* 133: 1563–1576

Normand P. and Bousquet J. 1989 Phylogeny of nitrogenase sequences in *Frankia* and other nitrogen-fixing microorganisms. *J. Mol. Evol.* 29: 436–447

Normand P., Simonet P. and Bardin R. 1988 Conservation of *nif* sequences in *Frankia*. *Mol. Gen. Genet.* 213: 238–246

Osawa S., Jukes T. H., Muto A., Yammo F., Ohama T. and Andachi Y. 1987 Role of directional mutation pressure in the evolution of the eubacterial genetic code. *Cold Spring Harbor Symp. Quant. Biol.* 52: 777–789

Pretorius I. M., Rawlings D. E., ONiell E. G., Jones W. A. and Woods D. R. 1987 Nucleotide sequences of the gene encoding the nitrogenase iron protein of *Thiobacillus ferrooxidans*. *J. Bacteriol.* 169: 367–370

Quinto C., de La Vega H., Flores M., Leemans J., Cevallos M. A., Pardo M. A., Azpiroz R., de Lourdes Girard M., Calva E. and Palacios R. 1985 Nitrogenase reductase: a functional multigene family in *Rhizobium phaseoli*. *Proc. Natl. Acad. Sci. USA* 82: 1170–1174

Raina R., Reddy M. A., Ghosall D. and Das H. K. 1988 Characterization of the gene for the Fe-protein of the vanadium dependent alternative nitrogenase of *Azotobacter vinelandii* and construction of a Tn5 mutant. *Mol. Gen. Genet.* 214: 121–127

Robson R., Woodley P. and Jones R. 1986 Second gene *nif*H* coding for a nitrogenase iron protein in *Azotobacter chroococcum* is adjacent to a gene coding for a ferredoxin-like protein. *EMBO J.* 5: 1159–1163

Ruvkun G. B. and Ausubel F. M. 1980 Interspecies homology of nitrogenase genes. *Proc. Natl. Acad. Sci. USA* 77: 191–195

Schumann J. P., Waitches G. M. and Scolnik P. A. 1986 A DNA fragment hybridizing to a *nif* probe in *Rhodobacter capsulatus* is homologous to a 16S rRNA gene. *Gene* 48: 81–92

Scolnik P. A. and Haselkorn R. 1984 Activation of extra copies of genes coding for nitrogenase in *Rhodopseudomonas capsulata*. *Nature* (*London*) 307: 289–292

Scott K. F., Rolfe G. B. and Shine J. 1983a Nitrogenase structural genes are unlinked in the nonlegume symbioant *Parasponia rhizobium*. *DNA* 2: 141–148

Scott K. F., Rolfe G. B. and Shine J. 1983b Biological nitrogen fixation: primary structure of the *Rhizobium trifoli* iron protein gene. *DNA* 2: 149–155

Souillard N., Magot M., Possot O. and Sibold L. 1988 Nucleotide sequence of regions homologous to *nif*H nitrogenase Fe protein from the nitrogen-fixing archaebacteria *Methanococcus thermolithotrophicus* and *Methanobacterium ivanovii*: evolutionary implications. *J. Mol. Evol.* 27: 65–76

Souillard N. and Sibold L. 1986 Primary structure and expression of a gene homologous to *nif*H nitrogenase Fe protein from the archaebacterium *Methanococcus voltae. Mol. Gen. Genet.* 203: 21–28

Souillard N. and Sibold L. 1989 Primary structure, functional organization and expression of nitrogenase structural genes of the thermophilic archaebacterium *Methanococcus thermolithotrophicus. Mol. Microbiol.* 34: 541–551

Sundaresan V. K. and Ausubel F. M. 1981 Nucleotide sequence of the gene coding for nitrogenase iron protein from *Klebsiella pneumoniae. J. Biol. Chem.* 256: 2808–2812

Toerock I. and Kondorosi A. 1981 Nucleotide sequence of the *R. meliloti* nitrogenase reductase *nif*H gene. *Nucleic Acids Res.* 9: 5711–5723

Wang S. Z., Chen J. S. and Johnson J. K. 1987 Nucleotide and deduced amino acid sequences of *nif*D encoding the alpha-subunit of nitrogenase Mo-Fe protein of *Clostridium pasteurianum. Nucleic Acids Res.* 15: 3935

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

Weinmann J. J., Fellows F. F., Gresshoff P. M., Shine J. and Scott K. F. 1984 Structural analysis of the genes encoding the molybdenum-iron protein of nitrogenase in the *Parasponia rhizobium* strain ANU289. *Nucleic Acids Res.* 12: 8329–8344

Wilbur W. J. and Lipman D. J. 1983 Rapid similarity searches of nucleic acid and protein data banks. *Proc. Natl. Acad. Sci. USA* 80: 726–730

Woese C. R. 1987 Bacterial evolution. *Microbiol. Rev.* 221–271

Yunn A. C. and Szalay A. A. 1984 Structural genes of dinitrogenase and dinitrogenase reductase are transcribed from two separate promoters in the broad host range cowpea *Rhizobium* strain *Irc78. Proc. Natl. Acad. Sci. USA* 81: 7358–7362