Functional divergence outlines the evolution of novel protein function in NifH/BchL protein family

SUBARNA THAKUR[1], ASIM K BOTHRA[2], and ARNAB SEN[1,*]

[1]*NBU Bioinformatics Facility, Department of Botany, University of North Bengal, Siliguri, India*

[2]*Bioinformatics Chemoinformatics Laboratory, Department of Chemistry, Raiganj College, Raiganj, India*

**Corresponding author (Email, senarnab_nbu@hotmail.com)*

Biological nitrogen fixation is accomplished by prokaryotes through the catalytic action of complex metalloenzyme, nitrogenase. Nitrogenase is a two-protein component system comprising MoFe protein (NifD&K) and Fe protein (NifH). NifH shares structural and mechanistic similarities as well as evolutionary relationships with light-independent protochlorophyllide reductase (BchL), a photosynthesis-related metalloenzyme belonging to the same protein family. We performed a comprehensive bioinformatics analysis of the NifH/BchL family in order to elucidate the intrinsic functional diversity and the underlying evolutionary mechanism among the members. To analyse functional divergence in the NifH/BchL family, we have conducted pair-wise estimation in altered evolutionary rates between the member proteins. We identified a number of vital amino acid sites which contribute to predicted functional diversity. We have also made use of the maximum likelihood tests for detection of positive selection at the amino acid level followed by the structure-based phylogenetic approach to draw conclusion on the ancient lineage and novel characterization of the NifH/BchL protein family. Our investigation provides ample support to the fact that NifH protein and BchL share robust structural similarities and have probably deviated from a common ancestor followed by divergence in functional properties possibly due to gene duplication

[Thakur S, Bothra AK and Sen A 2013 Functional divergence outlines the evolution of novel protein function in NifH/BchL protein family. *J. Biosci.* **38** 733–740] **DOI** 10.1007/s12038-013-9360-5

1. Introduction

Atmospheric nitrogen fixation carried out by bacteria is an important facet of global nitrogen balance. This ability is found in almost all major groups of bacteria including few methanogenic archaea and is carried out by an enzyme system called nitrogenase. The nitrogenase system is a complex metalloenzyme with conserved structural and mechanistic features. The most widespread form of this enzyme is a two-component system that consists of an iron protein (component 2) encoded by *nif*H and molybdenum-iron protein (component 1) composed of two non-identical subunits, α and β, working in tandem for nitrogen reduction (Burgess and Lowe 1996; Rees 2002; Rees *et al.* 2005). The α and β subunits are encoded by the *nif*D and *nif*K genes respectively. Besides the conventional molybdenum-based nitrogenase, there also exists alternative nitrogenases that contain vanadium instead of molybdenum in their component 1, encoded by *vnf* genes, and another form that contain only iron, encoded by *anf* genes (Bishop and Premakumar 1992) . However, sequence-wise they share similarity and probably arose from a common ancestor (Bothe *et al.* 2007). The NifH protein unit also have a large number of distant relatives like light-independent protochlorophyllide (Pchlide) reductase (ChlL/FrxC or BchL), which is involved in light-independent chlorophyll biosynthesis in anoxygenic photosynthetic bacteria and MinD that functions in spatial regulation of cell division (de Boer *et al.* 1992; Burke *et al.* 1993). These proteins are apparently quite distant in their biological functions but they bear considerable sequence and structural similarities to the Fe-protein. In fact, because of the structural similarities, the ChlL/BchL and NifH proteins have been grouped into the same protein family. This provides an indirect support to the hypothesis that diazotrophy being an anaerobic ability had probably emerged from the photosynthesis coupled with a loss of functionality in non-selective environment (Postgate 1982;

Keywords. Divergence; metalloenzyme; NifH; phylogenetic; positive selection

Published online: 1 October 2013

Postgate and Eady 1988). There is also another view which states that nitrogenase probably did not evolve from protochlorophyllide reductase, but both had a common ancestor (Reinbothe *et al*. 2010). The shift in protein function from that of its ancestor could be a result of functional divergence. The gene duplication event within a gene family often leads to such functional divergence, and as a consequence, some residues encounter altered functional constraints (Landgraf *et al.* 1999; Dermitzakis and Clark 2001). The evolutionary rates at these sites will vary in different homologous genes of a gene family and this is known as Type 1 functional divergence (Landgraf *et al.* 1999; Dermitzakis and Clark 2001). Although the exact trajectory of evolution of diazotrophy is still unclear, the analysis of functional divergence of the enzymes involved in these processes may lead to better understanding of whole course. Taking this into account, in this study we looked into the possibility of functional divergence in the NifH/BchL family. Subsequently, amino acid sites that are involved in functional divergence in this family have also been identified. We also looked into whether positive Darwinian selection at the amino acid level had a role to play in this functional divergence. Finally, a 3D structure-based approach was applied to establish the ancient divergence and explore phylogenetic relations of novel proteins within this complex family. In this approach, pair-wise superpositions of 3D structures were used to calculate similarity scores as an input for a tree-building algorithm.

2. Materials and methods

2.1 *Data collection and multiple sequence alignments*

Amino acid and nucleotide sequences of NifH protein from 40 different diazotroph were obtained from JGI-IMG database (*www.img.jgi.doe.gov*) (Markowitz *et al.* 2006).The alternative nitrogenase sequences were not considered for this study since they have a sporadic occurrence across the microbial genome. Protein BLAST was carried out to identify potential homologues of the NifH from light-independent protochlorophyllide (Pchlide) reductase. A protein was considered as homologue when the amino acid identity was above 35% over a stretch of ≥150 amino acids. Five such sequences were selected for further analysis. A total of 45 protein sequences were subjected to multiple alignments using ClustalW 1.83 (Thompson *et al.* 1994). A phylogenetic tree was constructed based on neighbor-joining (NJ) method using the software MEGA 4 (Tamura *et al.* 2007).

2.2 *Functional divergence analysis*

Gene duplication events in the protein family were tested for Type 1 functional divergence utilizing the Diverge version 2.0 software which is based on the method developed by Gu (1999, 2001). This method is based on maximum likelihood procedures to estimate functional divergence between member genes of a protein family, based on site-specific change in evolutionary rates after gene duplication or speciation. If the site-specific change is significant, it is desirable to identify the amino acid residues experiencing a shift in functional constraints. These amino acids sites are often conserved in one subfamily but highly variable in another. Previously created NJ tree was utilized in this analysis and subclusters were pairwise compared to each other. The coefficient of functional divergence (θ) and the posterior probability for the functional divergence were calculated for each position in the alignment. The cut-off value for the posterior probability was determined by consecutively eliminating the highest scoring residues from the alignment until the coefficient of functional divergence dropped to zero. The site-specific profile was used to predict critical amino acid sites for functional divergence. The position of predicted functionally divergent residues were mapped on to the structure of NifH protein and solubility accessibility of the residues were assessed by the ASAView server (Ahmad *et al.* 2004)

2.3 *Detection of positive selection*

DNA sequences along their relevant protein sequence multiple alignments were submitted to the PAL2NAL web server (Suyama *et al.* 2006), which provide a corresponding codon alignment for the sequences. Based on protein alignment maximum likelihood (ML) phylogenetic tree was reconstructed using PhyML software package (Guindon and Gascuel 2003). ProtTest (Abascal *et al.* 2005) was used for searching best fit model of protein evolution, and subsequently JTT+I+G model was selected for tree construction. The codon alignment and ML tree were further provided to CODEML package (Yang 1997) and the site-specific models M7andM8 were tested using the likelihood ratio test. These models examines variation in ω (dN/dS) among sites The Bayes empirical approach was employed to identify positively selected sites with a good confidence under the likelihood framework.

2.4 *Construction of structure-based phylogenetic tree*

The DALI algorithm (Holm and Sander 1993) was used for the pair-wise comparison and alignment of 3D structures of NifH protein and its relative structural neighbours. The DALI algorithm was utilized for calculating the root mean square distances, i.e. RMSD (the measure of the average distance between the backbone atoms of superimposed proteins), which serve as indicators of protein similarity. The resulting RMSD values were employed for construction of distance matrix. The matrix was used to create phylogenetic

(similarity) tree using the Fitch-Margoliash algorithm (Fitch and Margoliash 1967) as implemented in the FITCH program of the PHYLIP package (Felsenstein 1989).

3. Results

3.1 *Analysis of functional divergence*

A neighbour-joining-type phylogenetic tree based on the amino acid sequences was produced to assess the evolution of various members of the NifH/BchL family (figure 1). Here the NifH sequences form a clearly defined clade, but there are also a large number of more distant relatives; these include proteins involved in the synthesis of photosynthetic pigments, namely, protochlorophyllide reductase (BchL or ChlL) and chlorin reductase (BchX). Based on clustering, the phylogenetic tree has been further divided into six major subclusters of NifH and one of BchL sequences. NifH1 subcluster comprises NifH sequences from Cyanobacteria except one from *Anabaena*, whereas NifH2 subcluster has sequences from nitrogen-fixing actinobacteria like *Frankia*. NifH3 and 4 subclusters include legume-associated symbiotic bacteria along with *Beijerinckia* and *Azospirillum*, which are crop rhizospheric bacteria. Subcluster 5 and 6 consist of free-living diazotrophs and anaerobic clostridia group. BchL subcluster includes sequences of protochlorophyllide reductase from various anoxygenic phototrophic bacteria. This phylogeny was further used to detect site-specific altered functional constraints by comparing variability in evolutionary rates.

Table 1 shows the coefficient of functional divergence (θ) in various pair-wise comparisons. In all the cases, $\theta>0$ with $p<0.05$ indicating that likelihood ratio test implemented in detecting variability in evolutionary rates is found to be significant. It points towards rejection of null hypothesis, i.e. $\theta=0$ signifying that evolutionary rate of the amino acid residues has significantly shifted between gene clusters. Gene duplication could be the possible reason behind this site-specific rate shift in the protein family. Further analysis was subsequently focused on pair-wise comparison of the BchL/NifH2 and BchL/NifH5 subclusters. We had chosen these subclusters as they include members of an ancient symbiotic diazotroph and those in free-living condition. Amino acids residues responsible for functional divergence were identified using site-specific profiles (figure 2a and b) in combination with suitable cut-off values derived from the posterior probability of each comparison. The cut-off value for the posterior probability was determined by consecutively eliminating the highest scoring residues until the coefficient dropped to zero. Residues predicted to be functionally divergent were mapped onto the 3D model of nitrogenase iron protein (figure 3a and b).

In the BchL/NifH2 comparison, positions of the functionally divergent residues are: 100, 102,104, 105, 109,111, 113, 122,136, 154, 156, 157, 161, 168, 199, 227, 234, 248, 252, 253, 257, 259, 290, 293, 296 and 325. Eight residues (having position 102, 104, 109, 136, 168, 224, 257 and 296) were found to be functionally divergent while comparing between BchL and NifH5 subclusters. We further looked into the arrangement of these residues in the 3D structure of NifH protein with relevance to their solvent accessibility. The spiral plot (figure 4) reveals the position of the residues according to the accessible surface area. From the plot, it can be seen that the functionally divergent residues, particularly those near the position 100, 102, 105, 111, 136, 156, 168, 234 and 259, are situated in the inner ring of the spiral. These residues are actually buried residues with relatively low solvent accessibility.

3.2 *Selective pressure among the amino acid sites in the NifH family*

The presence of positive selection at amino acid residues was tested by implementing the site-specific models in CODEML program. Likelihood rate test (LRT) performed between model M7 (beta) and M8 (beta and ω) divulges that M8 model is significantly favoured. Positively selected sites were identified under M8 using a Bayesian method. Eighteen amino acid sites were inferred to be under the influence of positive selection. Amongst them the sites with high posterior probabilities, i.e. Bayes Empirical Bayes (BEB) analysis >0.95, are Met-28, Glu-73, Gly-76, Arg-79, Tyr-180, Lys-188, Glu-189, Cys-221, Glu-222, Glu-225, Glu-229 and Arg-232. Some of these residues particularly those near the C-terminus have been known to mediate inter-subunit interactions. The residues in the C-terminal region wrap around the body of the opposing subunit and enhance the overall stabilization of the NifH dimer (Schlessman *et al.* 1998).

3.3 *Structure based evolutionary relationships*

Sequence similarities between nitrogenase iron protein (NifH) and metallo proteins like BchL, MinD-1 and other chromosomal segregation proteins have led to the speculation of whether all these proteins have diverged from a single common ancestor or from one of the proteins amongst this group. In order to understand the evolutionary relationship between these proteins and trace their descent, a phylogram based on their 3D structures was generated. A detailed list of the various protein 3D structures considered for structure-based phylogenetic analysis is provided in table 2. The evolutionary tree based on the structure (figure 5) suggests that NifH protein and ChlL (3FWY) share maximum similarities and have probably diverged most recently compared to others. It also suggests that both NifH and ChlL could have evolved from a MinD-like ancestor (1HYQ), a bacterial

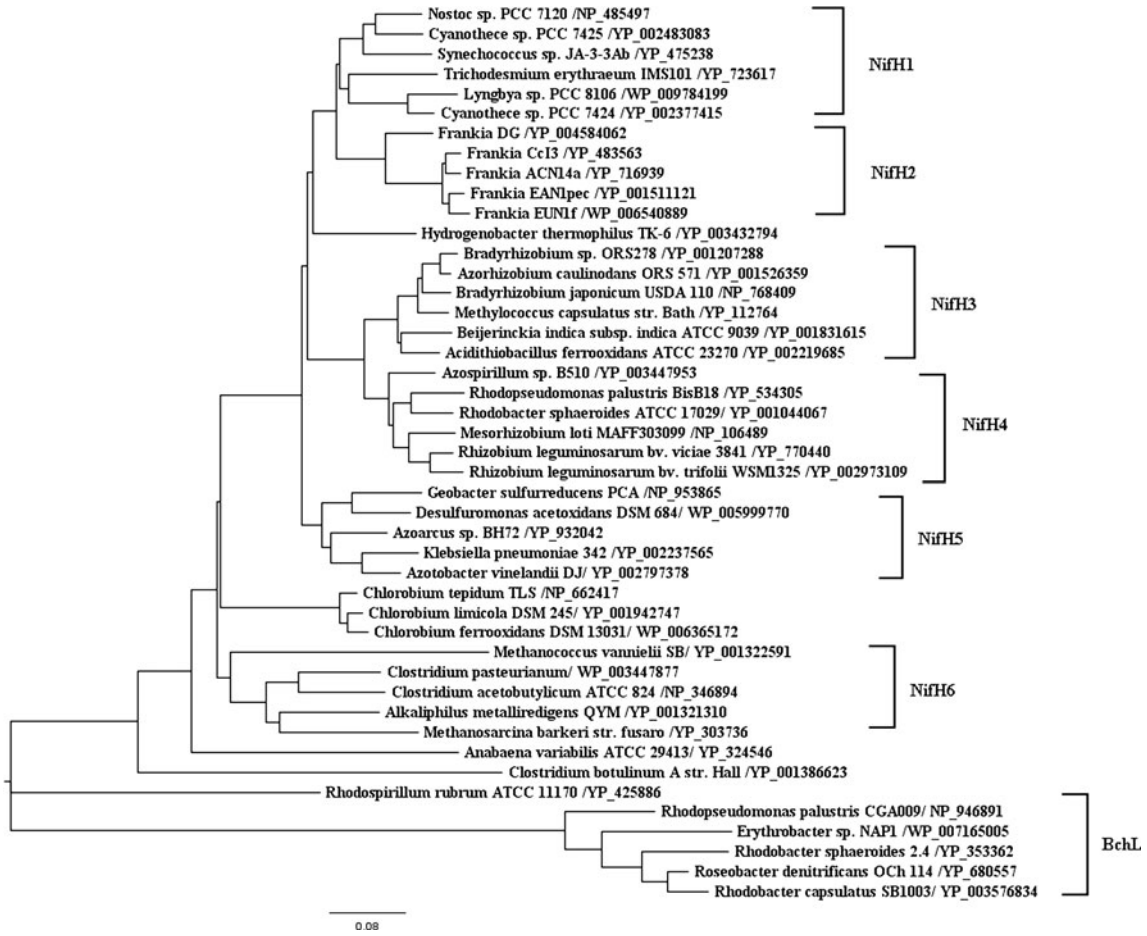

Figure 1. NJ-type phylogenetic tree amino acid sequence of the members of NifH/BchL protein family. The accession number of the sequences is mentioned along with the name of the organism.

cell division regulator protein which evolved parallelly with Soj chromosome segregation protein (2BEK). We also found that NifH proteins *Frankia* strains are relatively close to that of the cyanobacteria *Trichodesmium erythraeum*, while the structure of NifH of another cyanobacterium, *Anabaena variabilis*, is apparently quite apart and shows proximity with that of *Clostridium pasterium*. The NifH protein of *Bradyrhizobium* is structurally quite closer to that of *Azotobacter vinelandii* (1NIP and 1G5P).

4. Discussion

The phylogeny of NifH/BchL protein family has been utilized to detect the site-specific differential evolutionary change. Type 1 functional divergence analysis reveals that *bchL*/*nif* genes are significantly divergent functionally from each other, owing to the evolutionary rate and/or property differences at some amino acid sites. Gene duplication from a common ancestor might be the reason for the shift in the protein function. There have been other reports of in-tandem gene duplication in *nif*D and *nif*K giving rise to the functional components of the enzyme (Fani *et al.* 2000). Besides this, alternative supplemental forms of nitrogenase also been thought to have developed by gene duplications (Boyd *et al.* 2011). Hence, it seems our observation regarding functional divergence due to gene duplication in the NifH/BchL family is in line with that of earlier reports. The evolution of nitrogenase genes are indeed marked by gene duplications events. Further, in the study of functional divergence, amino acid residue relevant to the selective constrains has been identified based on a cut-off of functional divergence coefficient. These sites have varying evolutionary conservation among member genes. A closer look at their location in the protein structure in terms of accessible surface area revealed that they are actually buried residues which are involved in maintaining the correct 3D structural framework of the protein (figure 4).

Table 1. Estimates of the coefficient of functional divergence (θ) in various cluster pairs

Comparison among the subclusters	Theta ML	SE Theta	alphaML	LRT Theta
Bch1/NifH1	0.4408	0.1221	0.4418	13.0437
Bch1/NifH2	0.5712	0.1741	0.4188	10.7592
Bch1/NifH3	0.5632	0.1292	0.3242	18.9933
Bch1/NifH4	0.4656	0.1323	0.4045	12.3788
Bch1/NifH5	0.2768	0.1290	0.4351	4.6077
Bch1/NifH6	0.5656	0.1130	0.5985	25.0620
NifH1/NifH2	0.0960	0.1526	0.2887	0.3958
NifH1/NifH3	0.2216	0.0921	0.2327	5.7884
NifH1/NifH4	0.2056	0.1235	0.3048	2.7694
NifH1/NifH5	0.0744	0.0893	0.3089	0.6942
NifH1/NifH6	0.1592	0.0750	0.4775	4.5002
NifH2/NifH3	0.4917	0.1424	0.1576	11.9189
NifH2/NifH4	0.4944	0.1423	0.2255	12.0680
NifH2/NifH5	0.0952	0.1242	0.2588	0.5871
NifH2/NifH6	0.2032	0.1170	0.4952	3.0182
NifH3/NifH4	0.4720	0.1290	0.1916	13.3955
NifH3/NifH5	0.3376	0.1154	0.2124	8.5551
NifH3/NifH6	0.2856	0.1528	0.4451	3.4938
NifH4/NifH5	0.0584	0.1291	0.2847	0.2047
NifH4/NifH6	0.2352	0.1343	0.5115	3.0680
NifH5/NifH6	0.3048	0.1115	0.5097	7.4782

Theta ML represent the maximum likelihood estimate for theta, coefficient of functional divergence; AlphaML represent the maximum likelihood estimate for alpha, the gamma shape parameter for rate variation among sites; SE Theta represent standard error of the estimate theta and LRT Theta represent 2 log-likelihood-ratio against the null hypothesis of theta=0.

Functional divergence among proteins is often assumed to be strongly influenced by natural selection, particularly positive selection. Maximum likelihood (ML) tests implemented for investigating selective pressure detected positive selection in amino acids involved in inter-subunit interactions in NifH/BchL family. This indicates that positive selection has affected the evolution of these proteins. Therefore, to substantiate our prediction regarding the ancient divergence in this family, we then carried out phylogenetic analysis based on the 3D structures. For proteins sharing relatively low sequence identity as in this case, 3D structures are better than primary sequences for modelling of protein evolution (Balaji

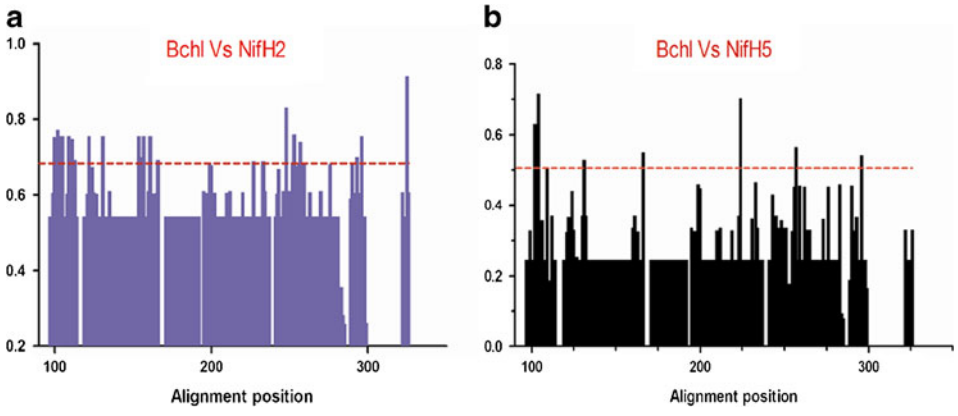

Figure 2. (**a**) and (**b**) Site specific profiles for evolutionary rate changes in the NifH/BchLprotein family. The posterior probabilities of functional divergence for NifH2, NifH5 and BchL were obtained with Diverge. Individual cut-off values for each comparison are marked with red horizontal lines.

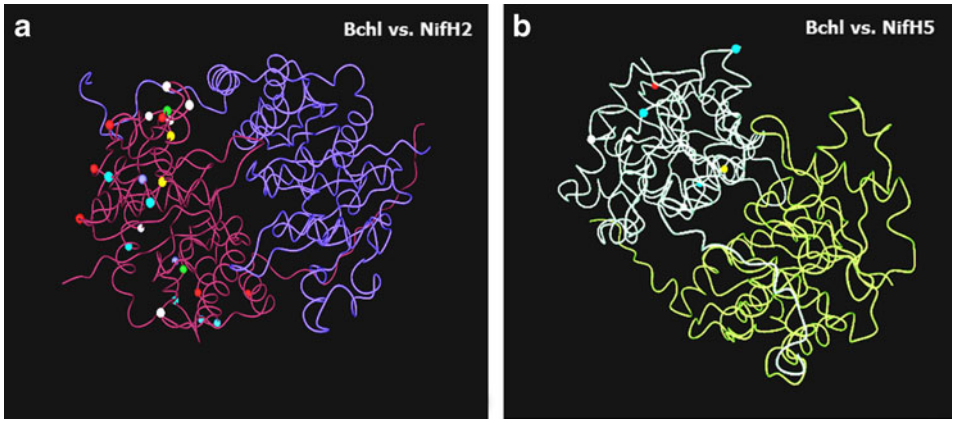

Figure 3. (**a**) and (**b**) Residues with predicted functional divergence between NifH/BchL subclusters are mapped onto 3D structure. The highlighted residues represent the position of functionally divergent sites.

and Srinivasan 2007). Results of the 3D-structure-based phylogenetic analysis provide evidences that both the NifH protein as well as the BchL protein have evolved from a common ancestor which is a MinD-like bacterial cell division inhibitor protein but during the course of evolution the photosynthesis-related protochlorophyllide reductase proteins have diverged from the main stock, followed by the NifH proteins.

5. Conclusion

The structure and sequence similarities between the nitrogenase iron proteins with those of proteins involved in photosynthesis has always been a matter of much curiosity. This study attempted a comprehensive approach to address the issue of molecular evolution of the NifH/BchL protein and

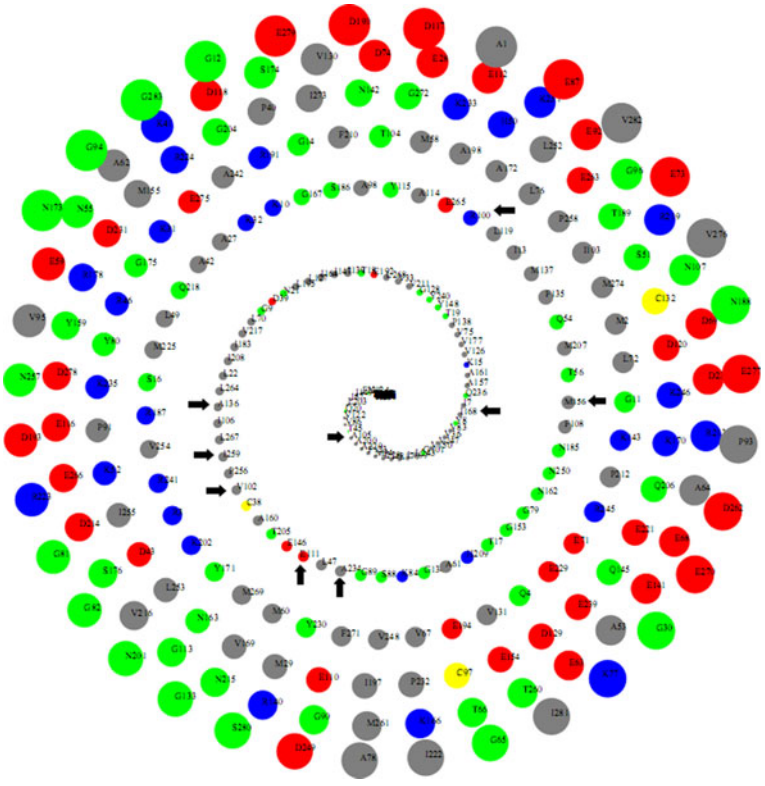

Figure 4. Solvent accessibility plot of the monomer of NifH protein. The black arrows represent the position of the functionally divergent residues with low solvent accessibility.

Table 2. Details of the NifH/BchL family 3D protein structures used for phylogenetic analysis

ID	Details about the protein	Strain
1NIP (PDB id)	Nitrogenase iron protein	*Azotobacter vinelandii*
1G5P (PDB id)	Nitrogenase iron protein (All-ferrous [4Fe-4S]0 form)	*Azotobacter vinelandii*
1CP2 (PDB id)	Nitrogenase iron protein	*Clostridium pasteurianum*
1G1M (PDB id)	Nitrogenase iron protein (All-ferrous [4Fe-4S]0 form)	*Azotobacter vinelandii*
2AFH (PDB id)	Nitrogenase complex	*Azotobacter vinelandii*
3NFP; 3NHP; 3NGP	Nitrogenase iron protein# (Sen *et al.* 2010)	*Frankia* ACN14a, *Frankia* CcI3 and *Frankia* Ean1pec
TRI	Nitrogenase iron protein# (Zehr *et al.* 1997)	*Trichodesmium* spp.
BRA	Nitrogenase iron protein from# (Thakur *et al.* 2012)	*Bradyrhizobium* sp. ORS278
1ANA	Nitrogenase iron protein†	*Anabaena variabilis*
3FWY (PDB id)	L -protein of light-independent protochlorophyllide reductase (BchL)	*Rhodobacter sphaeroides*
1HYQ (PDB id)	Crystal structure of bacterial cell division regulator MinD.	*Archaeoglobus fulgidus*
2BEK (PDB id)	Bacterial chromosome segregation protein Soj	*Thermus thermophilus*

'#' stands for validly published protein Homology Model; '†' stands for Homology model id Q3M4L0 taken from SWISS-MODEL repository.

looked into the events leading to the functional distinction between the proteins. Type 1 functional divergence analysis and structural evaluations of the proteins lend support to the fact that NifH and BchL proteins perhaps share a common ancestor but further functional evolution are marked by altered site-specific selective constrains possibly due to gene duplication. Critical amino acid residues relevant for distinct functional properties have also been identified. Positive selection pressure on amino acid residues involved in subunit interactions have also been found to play a potent role in the course of evolution. These findings have highlighted the functional evolution of NifH/BchL

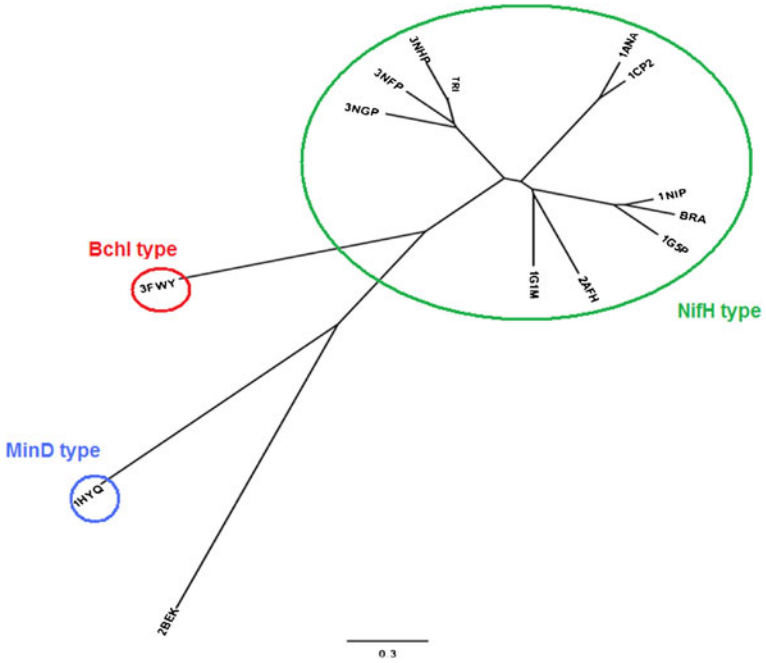

Figure 5. Structure-based phylogenetic tree of NifH/BchL family proteins inferred from the root mean square distances.

protein family and offer a potential starting point for further experimental verifications.

Acknowledgements

The authors are grateful to the Department of Biotechnology, India, for providing financial help in setting up Bioinformatics Centre, in the Department of Botany, University of North Bengal. ST would like to thank the CSIR for providing the CSIR-SRF research fellowship.

References

Abascal F, Zardoya R and Posada D 2005 ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* **21** 2104–2105

Ahmad S, Gromiha M, Fawareh H and Sarai A 2004 ASAView: database and tool for solvent accessibility representation in proteins. *BMC Bioinform.* **5** 51

Balaji S and Srinivasan N 2007 Comparison of sequence-based and structure-based phylogenetic trees of homologous proteins: Inferences on protein evolution. *J. Biosci.* **32** 83–96

Bishop PE and Premakumar R 1992 Alternative nitrogen fixation systems; in *Biological nitrogen fixation* (eds) G Stacey, RH Burris and HJ Evans (New York: Chapman and Hall) pp 736–762

Bothe H, Ferguson SJ and Newton WE 2007 *Biology of the nitrogen cycle* (Amsterdam; Boston: Elsevier)

Boyd ES, Hamilton TL and Peters JW 2011 An alternative path for the evolution of biological nitrogen fixation. *Front. Microbiol.* **2** 205

Burgess BK and Lowe DJ 1996 Mechanism of Molybdenum Nitrogenase. *Chem. Rev.* **96** 2983–3012

Burke DH, Hearst JE and Sidow A 1993 Early evolution of photosynthesis: clues from nitrogenase and chlorophyll iron proteins. *Proc. Natl. Acad. Sci. USA* **90** 7134–7138

de Boer PA, Crossley RE and Rothfield LI 1992 Roles of MinC and MinD in the site-specific septation block mediated by the MinCDE system of Escherichia coli. *J. Bacteriol.* **174** 63–70

Dermitzakis ET and Clark AG 2001 Differential selection after duplication in mammalian developmental genes. *Mol. Biol. Evol.* **18** 557–562

Fani R, Gallo R and 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 1989 PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics* **5** 164–166

Fitch WM and Margoliash E 1967 Construction of phylogenetic trees. *Science* **155** 279–284

Gu X 1999 Statistical methods for testing functional divergence after gene duplication. *Mol. Biol. Evol.* **16** 1664–1674

Gu X 2001 Maximum-likelihood approach for gene family evolution under functional divergence. *Mol. Biol. Evol.* **18** 453–464

Guindon S and Gascuel O 2003 A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52** 696–704

Holm L and Sander C 1993 Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233** 123–138

Landgraf R, Fischer D and Eisenberg D 1999 Analysis of heregulin symmetry by weighted evolutionary tracing. *Protein Eng.* **12** 943–951

Markowitz VM, Ivanova N, Palaniappan K, Szeto E, Korzeniewski F, Lykidis A, Anderson I, Mavromatis K, *et al.* 2006 An experimental metagenome data management and analysis system. *Bioinformatics* **22** e359–367

Postgate J and Eady R 1988 The evolution of biological nitrogen fixation; in *Nitrogen fixation: hundred years after* (eds) H Bothe, FJ de Bruijn, and WE Newton (Stuttgart: Gustav Fischer) pp 31–40

Postgate JR 1982 *The fundamentals of nitrogen fixation* (Cambridge: Cambridge Univ Press)

Rees DC 2002 Great metalloclusters in enzymology. *Annu. Rev. Biochem.* **71** 221–246

Rees DC, Akif Tezcan F, Haynes CA, Walton MY, Andrade S, Einsle O and Howard JB 2005 Structural basis of biological nitrogen fixation. *Philos. Transact. A. Math. Phys. Eng. Sci.* **363** 971–984; discussion 1035–1040

Reinbothe C, El Bakkouri M, Buhr F, Muraki N, Nomata J, Kurisu G, Fujita Y and Reinbothe S 2010 Chlorophyll biosynthesis: spotlight on protochlorophyllide reduction. *Trends Plant Sci.* **15** 614–624

Schlessman JL, Woo D, Joshua-Tor L, Howard JB and Rees DC 1998 Conformational variability in structures of the nitrogenase iron proteins from *Azotobacter vinelandii* and *Clostridium pasteurianum*. *J. Mol. Biol.* **280** 669–685

Sen A, Sur S, Tisa LS, Bothra AK, Thakur S and Mondal UK 2010 Homology modelling of the Frankia nitrogenase iron protein. *Symbiosis* **50** 37–44

Suyama M, Torrents D and Bork P 2006 PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **34** W609–612

Tamura K, Dudley J, Nei M and Kumar S 2007 MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24** 1596–1599

Thakur S, Bothra AK and Sen A 2012 *In silico* studies of NifH protein structure and its post-translational modification in Bradyrhizobium sp. ORS278. *Int. J. Pharm. Bio. Sci.* **3** B22 - B32

Thompson JD, Higgins DG and Gibson TJ 1994 CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22** 4673–4680

Yang Z 1997 PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13** 555–556

Zehr JP, Harris D, Dominic B and Salerno J 1997 Structural analysis of the Trichodesmium nitrogenase iron protein: implications for aerobic nitrogen fixation activity. *FEMS Microbial. Lett.* **153** 303–309