IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol. 9, Issue.1, January 2021 E-ISSN: 2347-2693

Data Mining and Phylogenetic Analysis of NifH Protein of *Azospirillum* strain among Nitrogen-fixing Bacteria using Bioinformatics Tools

Saurabh Sindhu[1*], Divya Sindhu[2], S.K. Yadav[3]

[1,2,3]Dept. of Computer Science and Engineering, Shri Jagdishprasad Jhabarmal Tibrewala University, Jhunjhunu - 333001, Rajasthan, India

**Corresponding Author: sindhu.saurabh04@gmail.com, Tel: +91-94166-50057*

DOI: https://doi.org/10.26438/ijcse/v9i1.110 | Available online at: www.ijcseonline.org

Received: 28/Dec/2020, Accepted: 13/Jan/2021, Published: 31/Jan/2021

Abstract **-** Development of prediction tools for computational identification of nitrogen fixation (*nif*) genes and categorization of potential diazotrophs using high throughput sequence data has accelerated the research in the area of biological nitrogen fixation. The computational tools are recently being used for the annotation and phylogenetic analysis of *nif*H gene or NifH protein sequences in nitrogen-fixing bacteria. In this study, phylogenetic analysis of NifH protein using Maximum Likelihood method showed that amino acid sequences of *Azospirillum brasilense* showed more relatedness to *Rhodospirillum rubrum* and *Rhodobacter capsulatus*. Further, the amino acid sequences also showed similarity to nodule-forming bacteria *Rhizobium etli*, *Rhizobium leguminosarum* bv. *trifolii* and *Sinorhizobium meliloti*. *Azospirillum brasilense* was placed on the same clade along with *Rhodopseudomonas palustris, Methylobacterium nodulans, Gluconoacetobacter diazotrophicus* and *Zymomonas mobilis* based on NifH aminoacid sequences. On another branch, NifH amino acid sequences of *Azospirillum brasilense* showed relatedness to *Bradyrhizobium diazoefficiens, Azorhizobium claudinodans* and *Acidothiobacillus ferrooxidans*. However, amino acid sequences of free-living nitrogen-fixing bacteria *Klebsiella pneumoniae*, *Azotobacter vinelanii* and *Azotobacter chroococcum* were also placed separately on other branch. Interestingly, sequences of anaerobic bacteria *Clostridium pasteurianum, Desulfatomaculum reducens* and *Chlorobium limicola* were placed far apart. Besides this, NifH amino acid sequences of *Azospirillum brasilense* showed quite divergence from the sequences observed in *Paenibacillus durus* and *Roseiflexus castenholzii*. Thus, NifH amino acid sequences classified various nitrogen-fixing bacteria into different phylogenetic clusters.

Keywords **-** NifH protein, Amino acids, Phylogenetic analysis, *Azospirillum*, Nitrogen fixation, Bioinformatics

I. INTRODUCTION

Biological nitrogen fixation (BNF) is one of the most ancient enzyme-catalyzed reactions [1], in which inert atmospheric nitrogen (N_2) is reduced to ammonical form in the soil using the nitrogenase enzyme [2]. This process of N_2 fixation is unique to certain bacteria, cyanobacteria, *Frankia* and archaea [3]. Biological nitrogen fixation is estimated to contribute for 90 x 10^{12} - 140 x 10^{12} g of fixed nitrogen annually into the biosphere [4]. The unpredictable occurrence of this trait across taxonomic groups in natural ecosystems [5], combined with the challenge of experimental detection of nitrogen fixation in agricultural systems [6], makes it difficult to obtain a comprehensive census of microorganisms with the capacity for nitrogen fixation (diazotrophy). The use of such nitrogen-fixing bacteria as biofertilizers is an environmental friendly technology for achieving sustainable restoration of soil fertility [7].

All nitrogen-fixing microorganisms rely on the nitrogenase enzyme for nitrogen reduction, which consists of as many as 20 *nif* genes [8, 9]. The nitrogenase enzyme complex consists of main structural genes *nif*H, *nif*D and *nif*K, which code for two metallo-proteins viz., iron (Fe) protein and iron-molybdenum (FeMo) protein [10, 11]. The *nif*H gene, which codes for dinitrogenase reductase, plays critical role in transfer of electrons during nitrogen fixation process and therefore, has been used as a molecular marker for identification of highly diversified nitrogen-fixing microbes from diverse environments [1, 12].

For improving biological nitrogen fixation, computational and bioinformatics tools are being applied for the annotation and phylogenetic analysis of *nif*H gene and NifH protein sequences [13, 14]. Besides identification of various *nif* genes using biotechnological techniques, computational tools are also used to identify and categorize novel potential diazotrophs [15]. In view of above prospects, biological data mining of NifH protein sequences was made from NCBI GenBank using BLAST and FASTA, and its phylogenetic analysis was carried out to find its relatedness in different diazotrophic microorganisms [16]. Better understanding about the functioning of nitrogenase enzyme using computational approach will supplement the existing biotechnological efforts to improve nitrogen fixation process for sustainable agriculture.

II. RELATED WORK

Most of the earlier studies have focussed either on the annotation of *nif*H sequences or on the phylogenetic distribution of diazotrophs by using *nif*H sequences as markers [13]. The diversity of nitrogen-fixing microorganisms was reported to vary dramatically with different habitats selecting for different groups of nitrogen-fixing organisms [17, 18, 19]. Frank [11] presented a novel approach to classify NifH protein sequences into well-defined phylogenetic clusters that provide a common platform for cross-ecosystem comparative analysis. The utility of this novel sequence binning approach revealed a marine – terrestrial distinction in the community composition. Frank et al. [12] developed computational method based on classification and regression trees (CART) for the annotation of *nif*H gene sequences, where the *nif*H protein sequences were classified into different phylogenetic clusters.

Dos Santos et al. [15] identified 149 diazotrophic species from fully sequenced genomes, including 82 known diazotrophs and 67 species not known to fix nitrogen. Computational prediction of nitrogen fixation was reported by using a new criterion with minimum set of six genes, coding for structural and biosynthetic components of nitrogenase namely *nif*HDK and *nif*ENB. Gaby and Buckley [20] designed a database that contains 32954 aligned nitrogenase *nif*H sequences that facilitated phylogenetic and evolutionary studies of nitrogen-fixing microorganism. Further, a software pipeline ARBitrator, was developed for retrieving autocurated *nif*H sequences from GenBank [21]. Similarly, Lau et al. [22] carried out studies on community phylogenetics and phylogeography of microorganisms living in extreme environments. The taxonomic assignments or patterns may give limited inference on how microbial functions are affected by historical, geographical and environmental factors.

III. MATERIALS AND METHODS

Data mining of the availabile microbial genome and protein sequences affords novel opportunities to provide the analysts with novel and efficient computational tools that overcome the constraints posed by the traditional statistical methods. Likewise, bioinformatics has evolved tremendously in recent years due to the explosive growth of biological information generated by the scientific community [16]. Phylogeny and phylogenetic trees give a picture of the evolutionary history among species, individuals or genes [23]. The need to query biological data using sets of evolutionary related taxa has spawned the need to create databases than can serve as repositories of phylogenetic trees.

3.1. Retrieval of NifH protein sequences in different nitrogen-fixing bacteria

Phylogenetic trees were built by character based methods using Maximum Likelihood (ML) method [24] that derives trees to optimize the distribution of the actual data pattern for each character. The ML method uses standard statistical techniques for inferring probability distributions to particular possible phylogenetic trees and allows additional statistical flexibility by permitting varying rates of evolution across both lineages and sites. Thus, Maximum Likelihood is well suited to the analysis of distantly related sequences [25].

FASTA was the first database similarity search tool that work on heuristic method of database searching. It uses a "hashing" submission of a query sequence and performed sequence to sequence pairwise comparison of the query sequence with all individual sequences available in that database. BLAST was used for searching of GenBank and other sequence databases for sequence similarity and homology among different nitrogen-fixing bacteria [26]. To access GenBank and its related retrieval and analysis services, the NCBI homepage: www.ncbi.nlm.nih.gov. was used as the search point [27].

In the present study, amino acid sequences of NifH protein from 97 different nitrogen-fixing and nodule-forming bacterial strains were retrieved from NCBI GenBank (Figure 1). Phylogenetic analysis, with other nitrogen-fixing bacteria and nodule-forming rhizobial species, was done by taking NifH protein sequences of *Azospirillum brasilense*, which fixes nitrogen as associative symbiont in cereal crops.

3.2. Phylogenetic analysis of NifH protein among different nitrogen-fixing bacteria using computational software

Amino acid sequences of NifH protein were searched for different nitrogen-fixing and nodule-forming rhizobia from NCBI GenBank and UniprotKB databases. Partial sequences for protein were removed from retrieved sequence datasets. The filtered sequences for proteins were aligned and conserved region and region of dissimilarity were identified from multiple sequence alignment using iterative and HMM algorithms of CLUSTALW/ CLUSTAL Omega program and MEGA software. Molecular Evolutionary Genetics Analysis (MEGA) computer software (i.e., MEGA-X) was used for statistical analysis of molecular evolution and for constructing phylogenetic trees. Alignment of sequences was carried out using CLUSTALW. Based on the nucleotide sequence database similarity, the relatedness of different amino acid sequences of NifH protein were compared by making the phylogenetic trees. Consensus trees were constructed for all sequences by bootstrapped method using both softwares and the number of replications (iterations) used to construct the phylogenetic tree were taken as 1000. Phylogenetic trees were generated graphically by using FigTree program, which is designed to display summarized and annotated files generated from a variety of programs, particularly those from BLAST output files. Generated trees were viewed using TREE VIEW and best fit tree was selected out of all trees.

IV. RESULTS

Among the nitrogen-fixing bacteria, *Azospirillum* species possess 'associative' lifestyle and these diazotrophs colonize the inner tissues of plants inter-cellularly. The plant growth stimulatory effect exerted by application of *Azospirillum* has been attributed to several mechanisms, including biological N_2 fixation [28]. Thus, understanding and optimizing these N_2-fixing plant-bacteria associations have promising prospective for sustainable agriculture [29]. Their application as biofertilizer for cereals will reduce the energy and pollution cost of the industrial reduction of N_2 from fertilizer industry. Besides this, the ecological impact and extent of N_2 fixation activity by making use of genetic engineering, biotechnological approaches, bioinformatics and computation modeling of proteins involved in nitrogen fixation is an interesting route for future investigations [12, 30, 31].

4.1. Sequence retrieval of amino acid sequences of NifH protein in nitrogen-fixing bacteria

In the present studies, NifH protein sequences from ninety seven different nitrogen-fixing bacterial strains and nodule-forming *Rhizobium* strains were retrieved from Uniprot KB Database (Figure 1). Amino acid sequences of NifH protein were further filtered out by ignoring partial, putative, hypothetical and uncharacterized sequences. The amino acid sequences for NifH protein was retrieved from NCBI GenBank with accession number P17303.1 in FASTA format. BLAST was used for searching of GenBank and other sequence databases for sequence similarity and homology among different nitrogen-fixing bacterial strains.

Twenty distinct amino acids are used by living cells for biosynthesis of various proteins. Amino acids combine to form long chains through formation of peptide bonds. The amino acid sequence makes up the primary structure of the protein. However, chemical/ biological properties of the protein are reliant on the 3 D or tertiary structure of the protein. Every amino acid has both a one-letter and three-letter abbreviation. Different amino acids are abbreviated as follows: Alanine (A), Arginine (R), Asparagine (N), Aspartic acid (D), Cysteine (C), Glutamine (Q), Glutamic acid (E), Glycine (G), Histidine (H), Isoleucine (I), Leucine (L), Lysine (K), Methionine (M), Phenylalanine (F), Proline (P), Serine (S), Threonine (T), Tryptophan (W), Tyrosine (Y) and Valine (V).

The amino acid sequences of NifH protein in different nitrogen-fixing organisms in FASTA format are provided. Although amino acid sequences of NifH protein were obtained from 97 nitrogen-fixing bacteria, the amino acid sequences of only eight microorganisms have been provided below. Following bacteria were included: *Azospirillum brasilense*, *Rhodobacter capsulatus*, *Sinorhizobium meliloti*, *Azotobacter chroococcum*, *Rhodobacter sphaeroides*, *Clostridium pasteurianum* and *Azorhizobium caulinodans.* These bacteria were isolated from differentat habitats and range from aerobic to anaerobic life style, free-living to nodule-forming symbiotic bacteria.

>*Azospirillum brasilense*

```
MSLRQIAFYGKGGIGKSTTSQNTLAALVELDQKILIVGCDPKADSTRLILHAKAQDTVLHLAAEAGSVEDLELEDVLKIGYK
GIKCVESGGPEPGVGCAGRGVITSINFLEENGAYDDVDYVSYDVLGDVVCGGFAMPIRENKAQEIYIVMSGEMMALYAANNI
AKGILKYAHSGGVRLGGLICNERQTDKEIDLASALAARLGTQLIHFVPRDNIVQHAELRRMTVIEYAPDSQQAQEYRQLANK
VHANKGKGTIPTPITMEELEEMLMDFGIMKSEEQQLAELQAKEAAKA
```

>*Rhodobacter capsulatus* SB 1003

```
MGKLRQIAFYGKGGIGKSTTSQNTLAALVEMGQKILIVGCDPKADSTRLILNTKLQDTVLHLAAEAGSVEDLEVEDVVKIGY
KGIKCTEAGGPEPGVGCAGRGVITAINFLEENGAYDDVDYVSYDVLGDVVCGGFAMPIRENKAQEIYIVMSGEMMALYAANN
IAKGILKYANSGGVRLGGLICNERKTDRELELAEALAAKLGCKMIHFVPRNNVVQHAELRRETVIQYDPTCSQAQEYRELAR
KIHENSGKGVIPTPITMEELEEMLMDFGIMQSEEDREKQIAEMEAAMKA
```

>*Sinorhizobium meliloti* 1021

```
MAALRQIAFYGKGGIGKSTTSQNTLAALVDLGQKILIVGCDPKADSTRLILNAKAQDTVLHLAATEGSVEDLELEDVLKVGY
RGIKCVESGGPEPGVGCAGRGVITSINFLEENGAYNDVDYVSYDVLGDVVCGGFAMPIRENKAQEIYIVMSGEMMALYAANN
IAKGILKYAHAGGVRLGGLICNERQTDRELDLAEALAARLNSKLIHFVPRDNIVQHAELRKMTVIQYAPNSKQAGEYRALAE
KIHANSGRGTVPTPITMEELEDMLLDFGIMKSDEQMLAELHAKEAKVIAPH
```

>*Azotobacter chroococcum* (strain mcd 1)

```
MAMRQCAIYGKGGIGKSTTTQNLVAALAEMGKKVMIVGCDPKADSTRLILHSKAQNTIMEMAAEAGTVEDLELEDVLKVGYG
GVKCVESGGPEPGVGCAGRGVITAINFLEEEGAYEDDLDFVFYDVLGDVVCGGFAMPIRENKAQEIYIVCSGEMMAMYAANN
ISKGIVKYANSGSVRLGGLCNSRNTDREDELIIALAAKLGTQMIHFVPRDNVVQRAEIRRMTVIEYDPTAKQADEYRTLARK
VVENKMLIIPNPITMDELEALLMEFGVMEEEDESIVGKAAAAEE
```


>*Rhodobacter sphaeroides*

MGKLRQIAFYGKGGIGKSTTSQNTLAALVEMGQKILIVGCDPKADSTRLILNTKLQDTVLHLAAEAGSVEDLELEDVVKIGY
KGIKCTEAGGPEPGVGCAGRGVITAINFLEENGAYDDVDYVSYDVLGDVVCGGFAMPIRENKAQEIYIVMSGEMMALYAANN
IAKGILKYANSGGVRLGGLICNERKTDRELELAEALAARLGCKMIHFVPRDNIVQHAELRRETVIQYAPESKQAQEYRELAR
KIHENSGKGVIPTPITMEELEEMLMDFGIMQSEEDRLAAIAAAEA

>*Clostridium pasteurianum*

MRQLAIYGKGGIAKSTTTQNLTAGLVERGNKIMVVGCDPKADSTRLLLGGLAQKTVLDTLREEGEDVELDSILKTGYAGIRC
VESGGPEPGVGCAGRGIITSINMLEQLGAYTDDLDFVFYDVLGDVVCGGFAMPIREGKAQEIYIVASGEMMALYAANNISKG
IQKYAKSGGVRLGGIICNSRKVANEYELLDAFAKELGSQLIHFVPRSPSVTKAEINKKTVIEYDPTCEQANEYRELARKVEE
NDMFVIPKPMTQERLEQILMEHGLID

>*Azorhizobium caulinodans* ORS 571

MSSLRQIAFYGKGGIGKSTTSQNTLAALAEMGHRILIVGCDPKADSTRLILHAKAQDTILSLAAAAGSVEDLELEEVMKIGY
RDIRCVESGGPEPGVGCAGRGVITSINFLEENGAYEDIDYVSYDVLGDVVCGGFAMPIRENKAQEIYIVMSGEMMAMYAANN
ISKGILKYANSGGVRLGGLVCNERQTDKELELAENLAKKLGTQLIYFVPRDNIVQHAELRRMTVIEYAPDSVQANHYRNLAE
RVHNNGGKGIIPTPITMDELEDMLMEHGIMKTVDETQVGKTAAELAALSA

4.2. Construction of phylogenetic trees based on sequence similarity index of NifH protein sequences

Phylogenetic trees are being used to navigate the sequences and to explore phylogenetic patterns found in associated metadata. The filtered amino acid sequences for NifH proteins, retrieved from UniprotKB databases, were aligned using CLUSTALW [32]. The amino acid sequences of NifH protein in different nitrogen-fixing bacteria was opened using MX: Alignment Explorer (Protein FASTA). The conserved regions and regions of dissimilarity were identified from multiple sequence alignment using iterative algorithms of CLUSTALW and MEGA software. Different amino acids showed consensus in different nitrogen-fixing bacteria. Conserved regions were found during the alignment at different positions of amino acids (Figure 2).

4.3. Phylogenetic tree construction

Multiple sequence alignments generated from CLUSTALW and MEGA software were used for generation of phylogenetic tree using MEGA-X software. Alignment of all retrieved sequences were done using CLUSTALW program. Consensus trees were constructed for all sequences by maximum likelihood method using MEGA-X. Generated trees were viewed using Figtree and best fit tree was selected out of all trees. Evolutionary analyses were conducted in MEGA-X. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site.

Amino acid sequences relatedness of NifH protein using Maximum Likelihood method showed that sequences *Azospirillum brasilense* showed amino acid similarity *Rhodospirillum rubrum* and *Rhodobacter capsulatus* (Figure 3). Further, the sequences also showed relatedness to nodule-forming bacteria i.e., *Rhizobium etli* (bean nodulating), *Rhizobium leguminosarum* bv. *trifolii* (clover nodulating bacteria) and *Sinorhizobium meliloti* (alfalafa nodulating). NifH aminoacid sequences were placed on the same clade along with *Rhodopseudomonas palustris, Methylobacterium nodulans, Gloconoacetobacter diazotrophicus* and *Zymomonas mobilis*. On the another branch, NifH amino acid sequences of *Azospirillum brasilense* showed relatedness to *Acidothiobacillus ferrooxidans, Bradyrhizobium diazoefficiens* and *Azorhizobium claudinodans*. The amino acid sequences of free-living nitrogen-fixing bacteria *Klebsiella pneumoniae*, *Azotobacter vinelanii* and *Azotobacter chroococcum* were also placed separately on other branch. Interestingly, sequences of anaerobic bacteria *Clostridium pasteurianum, Desulfatomaculum reducens* and *Chlorobium limicola* were placed far apart. NifH amino acid sequences of *Azospirillum brasilense* showed quite divergence from the amino acid sequences observed in *Paenibacillus durus* and *Roseiflexus castenholzii*.

Using the bootstrapping values in the Maximum Likelihood method, the phylogenetic reconstruction on a re-sampled set of data was performed. Amino acid sequences relatedness of NifH protein using Maximum Likelihood method showed that sequences *Azospirillum brasilense* showed amino acid similarity to *Rhodobacter spaeroids, Rhodobacter capsulatus* and *Rhodospirillum rubrum* (Figure 4). Further, the sequences also showed relatedness to nodule-forming bacteria i.e., *Azorhizobium caulinodans* (*Sesbania* nodulating), *Bradyrhizobium diazoefficiens, Herbaspirillum seropodicae* and *Acidothiobacillus ferrooxidans*. NifH aminoacid sequences were placed on another clade with *Rhodopseudomonas palustris, Zymomonas mobilis, Gloconoacetobacter diazotrophicus* and *Methylobacterium nodulans*. The amino acid sequences of free-living nitrogen-fixing bacteria *Klebsiella pneumoniae*, *Azotobacter vinelanii* and *Azotobacter chroococcum* were also placed separately on other branch. Sequences of anaerobic bacteria *Clostridium pasteurianum, Desulfatomaculum reducens, Chlorobaculum* sp. and *Chlorobium limicola* were placed far apart quite divergence from the NifH amino acid sequences of *Azospirillum brasilense.*

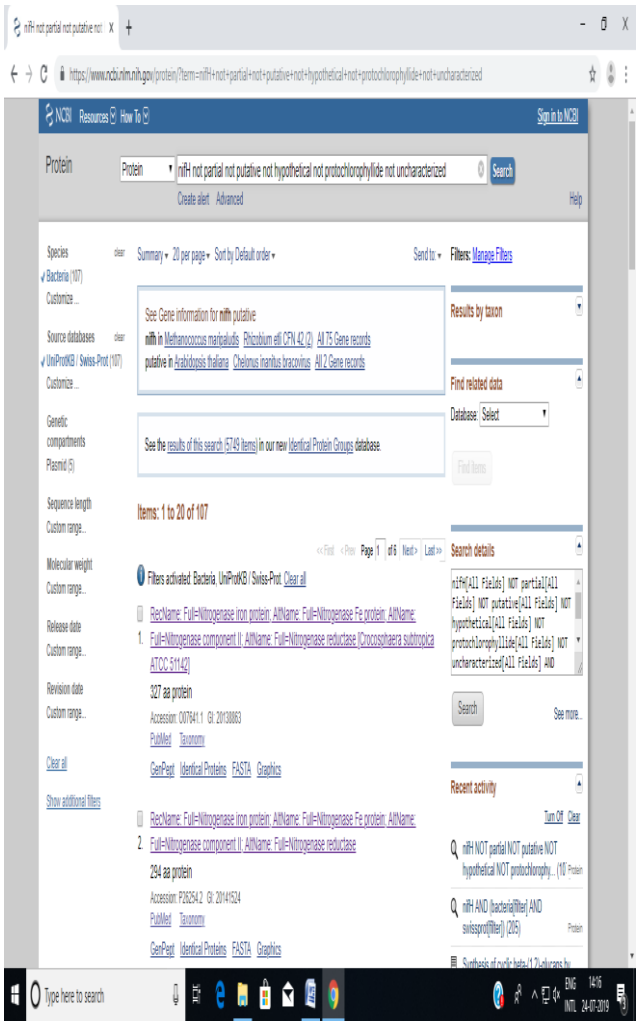

Figure 1. Sequence retrieval of NifH protein using query in NCBI GenBank

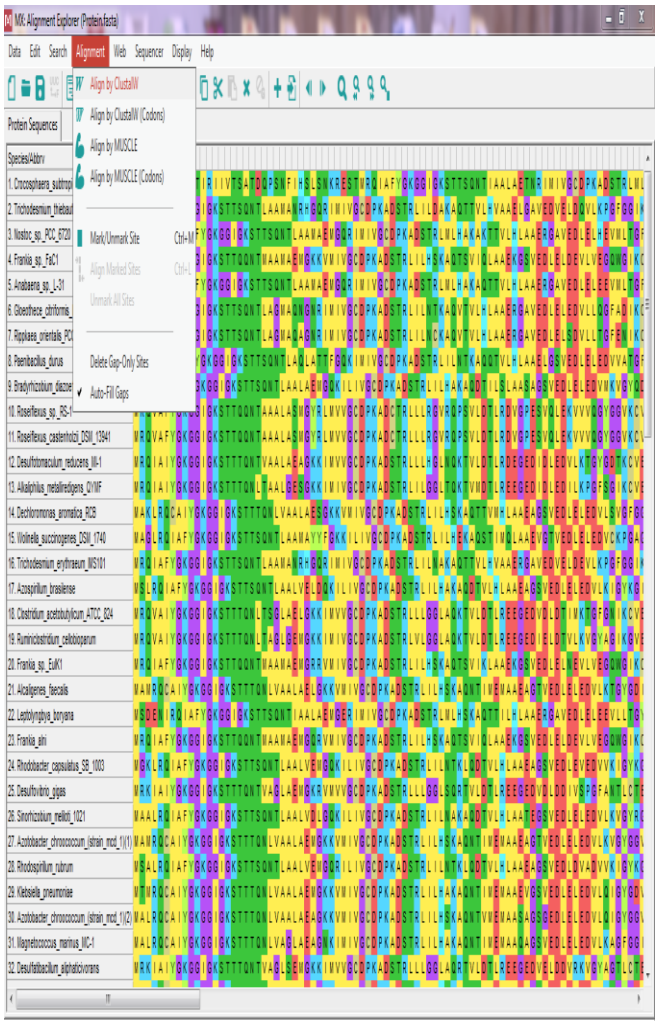

Figure 2. Interface of alignment of amino acid sequences of NifH protein in different nitrogen-fixing organisms using MEGA-X software

V. DISCUSSION

The free-living nitrogen-fixing organisms, containing nitrogenase enzyme complex, belong to the kingdoms eubacteria and archaebacteria. These bacteria supply fixed nitrogen to the global nitrogen cycle and therefore, diazotrophs are present in virtually all ecosystems, with representatives in environments as varied as aerobic soils (e.g., *Azotobacter* species), the ocean surface layer (*Trichodesmium*) and specialized nodules in legume roots (*Rhizobium*). The application of nitrogen-fixing *Azospirillum* strains as biofertilizers has been reported to enhance plant biomass and crop yield in a variety of crops worldwide [28, 29, 33, 34]. Considering the importance and application of nitrogenase enzyme in agricultural field, the present study was undertaken for phylogenetic analysis of NifH protein sequences in *Azospirillum brasilense*, which fixes nitrogen in association or within plant tissues.

In similar phylogenetic studies, Choo et al. [35] demonstrated clustering of *Paenibacillus azotofixans* NifH1 and NifH2 proteins within the *Cyanobacteriaceae* grouping. The NifH protein from a marine cyanobacterium, *Trichodesmium* sp. strain IMS101 showed the highest degrees of identity with *P. azotofixans* NifH1 (80%) and NifH2 (79%), respectively. The third putative *nif*H gene product of *P. azotofixans* (NifH3) clustered with NifH proteins of members of the *Archaea* domain, *Methanothermococcus thermolithotrophicus* and *Methanothermobacter thermoautotrophicus*. These phylogenetic analyses demonstrated that NifH1 and NifH2 form a monophyletic group among cyanobacterial NifH proteins, where as NifH3 sequences showed similarity with NifH proteins of the highly divergent methanogenic archaea.

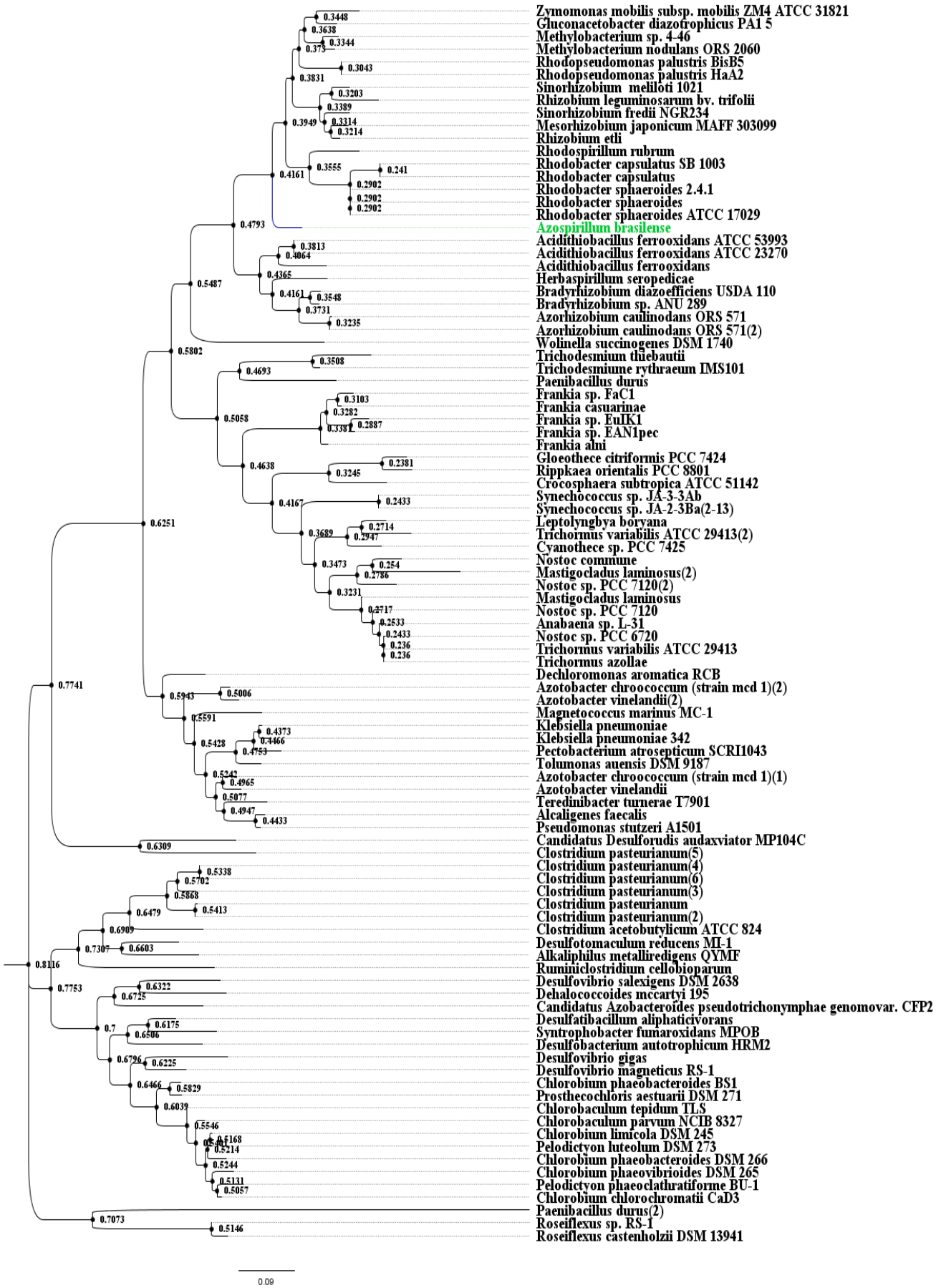

Figure 3. Phylogenetic tree of NifH protein by Maximum Likelihood method (without bootstrap) using MEGA-X

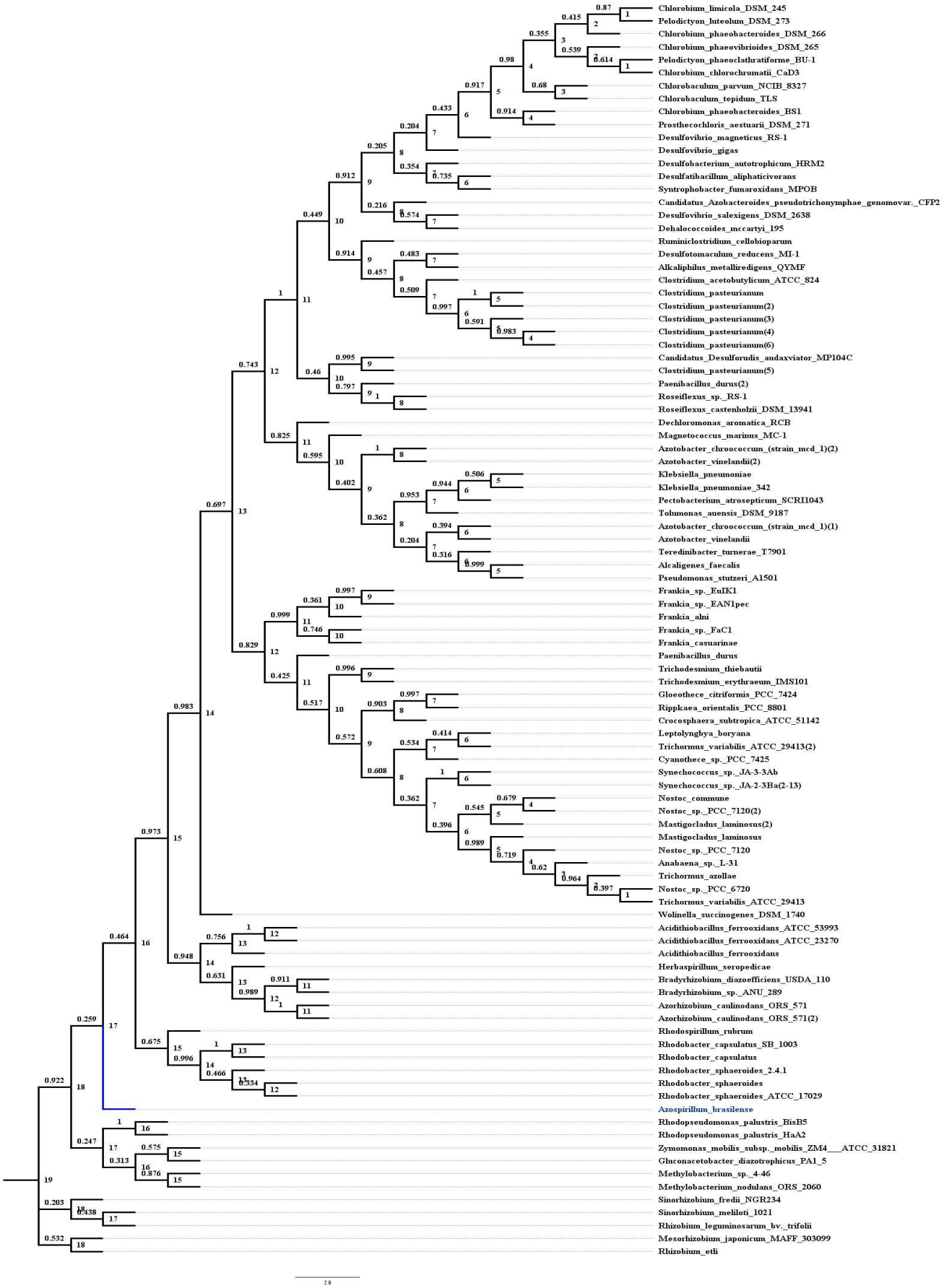

Figure 4. Phylogenetic tree of NifH protein using bootstrap method (Number of iterations = 1000) by Maximum Likelihood method in MEGA-X

Dos Santos et al. [15] suggested that nitrogenase emerged in anaerobes and later diversified into facultative anaerobes and aerobes. The transition of nitrogenase from anaerobic to facultative anaerobic and aerobic organisms was accompanied by a substantial increase in the number of *nif* genes i.e., from a minimum of 7 genes in the mesophilic archaeon *Methanocaldococcus* sp. strain FS406-22 to a *nif* gene cluster composed of 9–10 genes in facultative *Paenibacillus* species, with further increase to a maximum of 20 genes in obligate aerobes *Azotobacter vinelandii* [36]. Different diazotrophs were organized in the α-subgroup (species of e.g. *Gluconacetobacter*, *Azospirillum*), β-subgroup (species of e.g. *Derxia*, *Azoarcus*, *Burkholderia*, *Herbasprillum*), as well as in the γ-subgroup (species of e.g. *Azotobacter*, *Serratia*, *Klebsiella*, *Pseudomonas*, *Pantoea*). Rhizosphere inhabiting and nitrogen-fixing *Paenibacillus* species have been reported as exceptions that belong to the Firmicutes. On the other hand, *Azospirillum* species have a more 'associative' or 'facultative endophytic' lifestyle. Epiphytic diazotrophic bacteria like *Beijerinckia fluminensis* and *Azorhizophilus paspali* (*Azotobacter paspali*) are mainly isolated from rhizoplane [37, 38].

Due to increasing yield of *nif*H sequences from high-throughput technologies and increasing size of environmental sequence libraries, phylogenetic classification of *nif*H gene sequences requires a fast automated solution [12]. Gaby et al. [31] developed a computational pipeline to improve *nif*H sequence analysis, which infers taxonomy and optionally filters out paralog sequences. In addition, an empirical model was employed to derive optimal operational taxonomic unit (OTU) cutoffs for the *nif*H gene at the species, genus and family levels.

VII. CONCLUSION

Biological nitrogen fixation has significant impact on nitrogen cycles in diverse ecosystems and this process accounts for the majority of nitrogen transferred from the atmospheric reservoir into the biosphere. The distribution pattern of microbial communities suggests that nitrogen-fixing ability is evolutionary and mainly transmitted vertically with the widespread loss of function [39]. Besides use of nitrogen-fixing bacteria as biofertilizers to increase crop yield, image processing technology is also being applied for early detection and control of plant diseases leading to improved agricultural productivity [40]. In the present study, phylogeny of NifH protein of nitrogen-fixing bacteria *Azospirillum brasilense* was carried out. Ninety seven sequences of NifH protein sequences from different bacterial species were retrieved from NCBI for phylogenetic analysis. Amino acid sequences relatedness of NifH protein using Maximum Likelihood method showed that sequences of *A. brasilense* showed amino acid similarity with *Rhodospirillum rubrum* and *Rhodobacter capsulatus*. Further, the sequences also showed relatedness to *Rhizobium etli*, *Rhizobium leguminosarum* bv. *trifolii* and *Sinorhizobium meliloti*. On the another branch, NifH amino acid sequences of *Azospirillum brasilense* showed relatedness to *Acidothiobacillus ferrooxidans*, *Bradyrhizobium diazoefficiens* and *Azorhizobium claudinodans*. Interestingly, sequences of anaerobic bacteria *Clostridium pasteurianum*, *Desulfatomaculum reducens* and *Chlorobium limicola* were placed far apart. NifH amino acid sequences of *Azospirillum brasilense* showed quite divergence from the amino acid sequences of *Paenibacillus durus* and *Roseiflexus castenholzii*. The results obtained in the present studies demonstrate about the structural annotation of NifH protein sequences. In addition, this study would be very helpful to identify, verify and classify various type of NifH proteins found in nodulating and free-living nitrogen-fixing bacteria. Moreover, the correlation or phylogenetic relatedness could be assessed among the NifH protein sequences obtained from different taxa.

REFERENCES

[1] J. Raymond, J.L. Siefert, C.R. Staples et al., *"Environmental history of nitrogen fixation"*. Molecular Biology Evolution, Vol. **21**, pp. **541-554**, **2004**.

[2] J. Kim, D.C. Rees, *"Nitrogenase and biological nitrogen fixation"*, Biochemistry, Vol. **33**, pp. **389-397**, **1994**.

[3] J.P.W. Young, *"Phylogenetic classification of nitrogen-fixing organisms"*, in *Biological Nitrogen Fixation*, G. Stacey, R.H. Burris, H.J. Evans (eds.). New York: Chapman & Hall. pp. **43–86**, **1992**.

[4] P.M. Vitousek, J.D. Aber, R.W. Howarth, G.E. Likens, P.A. Matson, D.W. Schindler, W.H. Schlesinger, D.G. Tilman, "*Human alteration of global nitrogen cycle: sources and consequences"*. Ecology Applications, Vol. **7**, pp. **737–750**, **1997**.

[5] C.C. Cleveland, A.R. Townsend, D. S. Schimel, et al., *"Global patterns of terrestrial biological nitrogen (N_2) fixation in environmental ecosystems"*, Global Biogeochem. Cycles, Vol. **13**, pp. **623-645**, 1**999**.

[6] I.R. Kennedy, N. Islam, *"Current and potential contribution of asymbiotic nitrogen fixation to nitrogen requirements on farms: review"*, Australian Journal of Experimental Agriculture, Vol. **41**, pp. **447-457**, **2001**.

[7] B.B. Bohlool, J.K. Ladha, D.P. Garrity, T. George, *"Biological nitrogen fixation for sustainable agriculture: perspective"*, Plant and Soil, Vol. **141**, pp. **1-11**, **1992**.

[8] D.R. Dean, M.R. Jacobson, *"Biochemical genetics of nitrogenase"*, In: Biological Nitrogen Fixation, G. Stacey, R.H. Burris, H.J. Evans (eds.). Chapman-Hall, NewYork, pp. **763-831**, **1992**.

[9] Q. Li, X. Liu, H. Zhang, S. Chen, *"Evolution and functional analysis of orf1 within nif gene cluster from Paenibacillus graminis RSA19*", International Journal of Molecular Sciences, Vol. **20**, pp. **1145**, **2019**. doi:10.3390/ijms20051145

[10] T.L. Hamilton, M. Ludwig, R. Dixon, E.S. Boyd, P.C. Dos Santos, J. C. Setubal et al., *"Transcriptional profiling of nitrogen fixation in Azotobacter vinelandii"*, Journal of Bacteriology, Vol. **193**, pp. **4477–4486**, **2011**. doi: 10.1128/JB.05099-11

[11] I. Frank, *"Rapid classification of NifH protein sequences using classification and regression trees"*, M. Sc. dissertation, University of California, Santacruz, **2014**.

[12] I. Frank, K.A. Turk-Kubo, J. P. Zehr, *"Rapid annotation of nifH gene sequences using classification and regression trees facilitates environmental functional gene analysis"*, Environmental Microbiological Reports, Vol. **8**, No. **5**, pp. **905–916**, **2016**.

[13] U.K. Mondal, B. Das, T.C. Ghosh, A. Sen, A. K. Bothra, *"Nucleotide triplet based molecular phylogeny of class I and class II aminoacyl t-RNA synthetase in three domain of life process: bacteria, archaea and eukarya"*, Journal of Biomolecular Structure Dynamics, Vol. **26**, article no. **321328**, **2008**. doi: 10.1080/07391102.2008.10507247

[14] P.K. Meher, T.K. Sahu, J. Mohanty, S. Gahoi, S. Purru, M. Grover, A.R. Rao, *"nifPred: proteome-wide identification and categorization of nitrogen fixation proteins of diazotrophs based on composition-transition-distribution features using support vector machine",* Frontiers in Microbiology, Vol. **9**, pp. **1-16**, **2018**.

[15]P.C. Dos Santos, Z. Fang, S. Mason, J.C. Setuba, R. Dixon, *"Distribution of nitrogen fixation and nitrogenase-like sequences amongst microbial genomes",* BMC Genomics, Vol. **13**, pp. **162**, **2012**.

[16] M.A. Marti-Renom, A.C. Stuart, A. Fiser, R. Sanchez, F. Melo, A. Sali, *"Comparative protein structure modeling of genes and genomes",* Annual Review of Biophysics and Biomolecular Structures, Vol. **29**, pp. **291**, **2000**.

[17] J.P. Zehr, B.D. Jenkins, S.M. Short, G.F. Steward, *"Nitrogenase gene diversity and microbial community structure: cross–system comparison"*, Environonmental Microbiology, Vol. **5**, pp. **539-554**, **2003**.

[18]L.F.W. Roesch, F.A.O. Camargo, F.M. Bento, et al., *"Biodiversity of diazotrophic bacteria within the soil, root and stem of field-grown maize"*, Plant and Soil, Vol. **302**, pp. **91–104**, **2008**.

[19] J.C. Gaby, D.H. Buckley, *"Global census of nitrogenase diversity",* Environmental Microbiology, Vol. **13**, pp. **1790–1799**, **2011**.

[20] J.C. Gaby, D.H. Buckley, *"Comprehensive aligned nifH gene database: multipurpose tool for studies of nitrogen-fixing bacteria",* Database 2014:bau001, **2014**.

[21] P. Heller, H.J. Tripp, K. Turk-Kubo, J.P., Zehr, *"ARBitrator: a software pipeline for on-demand retrieval of auto-curated nifH sequences from GenBank",* Bioinformatics, Vol. **30**, pp. **2883–2890**, **2014**.

[22] M.C.Y. Lau, C. Cameron, C. Magnabosco, C.T. Brown, F. Schilkey, S. Grim, S. Hendrickson, M. Pullin et al., *"Phylogeny and phylogeography of functional genes shared among seven terrestrial subsurface metagenomes reveal N-cycling and microbial evolutionary relationships",* Frontiers in Microbiology, Vol. **5**, pp. **1-17**, **2014**.

[23]J.T.L. Wang, M.J. Zaki, H.T.T. Toivonen, D. Shasha, *"Data Mining in Bioinformatics",* Springer, **2005**.

[24]J. Felsenstein, *"Evolutionary trees from DNA sequences: Maximum likelihood approach",* Journal of Molecular Biology, Vol. **17**, pp. **368**, **1981**.

[25]B. Chor, T. Tuller, *"Maximum likelihood of evolutionary trees: hardness and approximation"*, Bioinformatics, Vol. **21** (Suppl 1), pp. **97-106**, **2005**.

[26]S.F. Altschul, T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, D. J. Lipman, *"Gapped BLAST and PSI-BLAST: new generation of protein database search programs",* Nucleic Acids Research**,** Vol. **25**, pp. **3389**, **1997**.

[27]Sayers, E.W., Barrett, T., Benson, D.A., Bryant, S.H., Canese, K., Chetvernin, V., Church, D.M., Dicuccio, M., Edgar, R., Federhen, S., et al., *"Database resources of National Center for Biotechnology Information"*, Nucleic Acids Research, Vol. **38**, pp. **D5–D16**, **2010**.

[28]Y. Okon, C.A. Labandera-Gonzalez, M. Lage, P. Lage, *"Agronomic applications of Azospirillum and other PGPR"*, In: Biological nitrogen fixation, F.J. de Brujin, (ed). Hoboken: Wiley, Chapter **90**, **2015**.

[29] Y. Okon, C.A. Labandera-Gonzalez, *"Agronomic applications of Azospirillum: evaluation of 20 years worldwide field inoculation",* Soil Biology and Biochemistry, Vol. **26**, pp. **1591–1601**, **1994**.

[30] A. Behar, B. Yuval, E. Jurkevitch, *"Enterobacteria- mediated nitrogen fixation in natural populations of the fruit fly Ceratitis capitata",* Molecular Ecology, Vol. **14, 2637-2643**, **2005**.

[31] J.C. Gaby, L. Rishishwar, L.C. Valderrama-Aguirre, S.J. Green, A. Valderrama-Aguirre, K. Jordan, J.E. Kostkaa, *"Diazotroph community characterization via high-throughput nifH amplicon sequencing and analysis pipeline",* Applied and Environmental Microbiology, Vol. **84**, Issue **4**, **e01512-17**, **2018**.

[32] D. Higgins, J.D. Thompson, D.G. Higgins, T.J. Gibson, *"CLUSTALW: Improving sensitivity of progressive multiple sequence alignment through sequence weighing, position-specific gap penalities and weight matrix choice"*, Nucleic Acids Research, Vol. **22**, pp. **4673-4680**, **1994**.

[33] Y. Bashan, G. Holguin, L.E. de-Bashan, *"Azospirillum-plant relationships: physiological, molecular, agricultural and environmental advances (1997–2003)*", Canadian Journal of Microbiology, Vol. **50**, pp. **521-577**, **2004**.

[34] L. Pereg, E. Luz, Y. Bashan, *"Assessment of affinity and specificity of Azospirillum for plants",* Plant and Soil, Vol. **399**, pp. **389–414**, **2016**.

[35] Q.-C. Choo, M.-R. Samian, N. Najimudin, *"Phylogeny and characterization of three nifH-homologous genes from Paenibacillus azotofixans",* Applied and Environmental Microbiology, Vol. **69**, No. **6**, pp. **3658–3662**, **2003**.

[36] E.S. Boyd, A.M.G. Costas, T.L. Hamilton, F. Mus, J.W. Peters, *"Evolution of molybdenum nitrogenase during transition from anaerobic to aerobic metabolism",* Journal of Bacteriology, Vol. **197**, pp. **1690–1699**, **2015**.

[37] B. Reinhold-Hurek, T. Hurek *"Life in grasses: diazotrophic endophytes",* Trends in Microbiology, Vol. **6**, pp. **139–144**, **1998**.

[38] J.I. Baldani, V.L.D. Baldani, *"History on biological nitrogen fixation research in graminaceous plants: special emphasis on Brazilian experience*", Annals of Academy Brasilean Ciênc, Vol. **77**, pp. **549–579**, **2005**.

[39] L.S. Hartmann, S.R. Barnum, *"Inferring evolutionary history of Mo-dependent nitrogen fixation from phylogenetic studies of nif*K *and nif*DK*",* Journal of Molecular Evolution, Vol. **71**, pp. **70-85**, **2010**

[40]D. Sindhu, S. Sindhu, *"Image processing technology application for early detection and classification of plant diseases"*, International Journal of Computer Science and Engineering, Vol. **7**, Issue **5**, pp. **92-97**, **2019**

AUTHORS PROFILE

Saurabh Sindhu received the B.C.A. and M.C.A. degrees from Guru Jambheshwar University of Science and Technology, Hisar. His areas of interest include data mining, data security, DNA computers, cloud computing, bioinformatics and computation modcling. He has published 23 research articles in national and international journals. He has participated in 17 national / international seminars and conferences. He is pursuing his Ph. D. degree in Department of Computer Science from Shri JJT University, Jhunjhunu, Rajasthan. He is having five years of teaching experience at the post of Assistant Professor in Department of Computer Science at CRM Jat College, Hisar.

Divya Sindhu received the B. Tech. degree in Computer Science and Engineering from Kurukshetra University, Kuruksheta in 2012. She successfully completed her M. Tech. degree in Computer Science and Engineering from Guru Jambheshwar University of Science and Technology, Hisar in 2015. Her areas of interest include computer algorithms, data mining, data security, cloud computing, computer networks, bioinformatics and computation modeling. She has published 25 research articles in national and international journals. She has participated in 18 national / international seminars and conferences. She is pursuing her Ph. D. degree in Department of Computer Science and Engineering from Shri JJT University, Jhunjhunu, Rajasthan. She is having five years of teaching experience at the post of Assistant Professor in Department of Computer Science at CRM Jat College, Hisar.

Dr. S. K. Yadav is presently serving as President of Shri JJT University, Jhunjhunu, Rajasthan. He has successfully completed his term as Honorary Secretary, Computer Society of India (2018-2020). He is Fellow member of the Institution of Electronics and Telecommunication Engineers, Indian Institute of Metals and All India Mangement Association. Dr. Yadav has 30 years of teaching and research experience in premiere institutions including University of Delhi. He has successfully guided more than 80 doctoral scholars with a credit of more than 150 research papers, 24 patents and 50 internationally renowned books on Computer Science, Mathematics, Education Management and Research. He had been Higher Education Ambassador of Bhutan, Nepal, Sri Lanka, Afghanistan and Indonesia for two terms.

