Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 02, February-2021

Computational Analysis of Phylogenetic Diversity and Evolutionary Relationships using *nif*H Gene Sequences Among Nitrogen-Fixing Organisms

Saurabh Sindhu[*1], Neeru Singh Redhu[2], Divya Sindhu[1] and S. K. Yadav[1]
[1]Department of Computer Science and Engineering
Shri Jagdishparsad Jhabarmal Tibrewala University, Jhunjhunu, Rajasthan, India
[2]Department of Molecular Biology, Biotechnology and Bioinformatics, CCS Haryana Agricultural University, Hisar – 125004, Haryana, India

***Abstract*:- Analysis of genes and proteins using in silico tools has been receiving greater attention currently to find suitable biomarkers for rapid identification of beneficial microbes. The use of nitrogen-fixing organisms as biofertilizer affords novel eco-friendly technology for increasing agricultural crop productivity. Recently, nitrogen fixation (nif) genes encoding nitrogenase proteins have been identified using DNA sequence analysis and transcriptional profiling. In this study, nifH gene sequences from 40 different nitrogen-fixing bacterial species were retrieved from NCBI GenBank for phylogenetic analysis. Phylogenetic trees were constructed using Maximm Likelihood method. It was found that nifH gene sequences of Azospirillum brasilense showed close similarity with Rhodobacter capsulatus and Rhodospirillum rubrum. Further, the nucleotide sequences also showed relatedness to nodule-forming bacteria i.e., Rhizobium leguminosarum, Sinorhizobium meliloti, Bradyrhizobium japonicum and Azorhizobium caulinodans. On another branch, nifH gene sequences of Azospirillum brasilense showed relatedness to Frankia sp. and also showed similarity with free-living nitrogen-fixing bacteria Klebsiella pneumoniae and Azotobacter vinelandii. The branching order of the nifH phylogeny showed that nitrogen-fixing bacteria isolated from diverse environments were distantly related. For example, Cyanobacterium sp., Trichodesmium erythraeum, Mastigocladus laminosus, Paenibacillus sabinae and Nostoc commune were placed on different branch showing large variations in their nucleotide sequences.**

Keywords – ***nifH gene, nucleotide sequences, Phylogenetic analysis, nitrogen fixation, Computational analysis***

1. INTRODUCTION

The process of nitrogen (N_2) fixation is sporadically distributed and unique to both eubacteria and methanogenic archaea [1, 2]. Many bacterial species including *Azotobacter*, *Azospirillum*, *Klebsiella*, *Rhizobium* and *Nostoc* have been found to reduce atmospheric inert nitrogen to plant utilizable ammonical form using the nitrogenase enzyme [3]. Terrestrial N_2 fixation is estimated to contribute for 90 x 10^{12} - 140 x 10^{12} g of fixed nitrogen per year [4] whereas, free-living and symbiotic N_2 fixation contribute in great part of fixed nitrogen in the N cycle and in agricultural systems [5, 6]. Among various nitrogen-fixing bacteria, *Azospirillum* species possess 'associative endophytic' lifestyle and these plant-associated diazotrophs colonize the inner tissues of plants inter-cellularly, without causing any apparent damage to the host [7]. The inoculation of nitrogen-fixing *Azospirillum* strains has been found to enhance plant growth, which has been attributed to several mechanisms, including biological N_2 fixation [8, 9]. Therefore, understanding and

optimization of these N_2-fixing plant-bacteria associations have promising prospective for sustainable agriculture.

The recent and rapid increase in the availability of microbial genome sequences using next-generation sequencing techniques [10] and informatics-based approaches provides novel opportunities to examine the occurrence and distribution of nitrogen fixation (*nif*) genes [11]. The detailed nucleotide sequence information of the various *nif* genes encoding the nitrogenase enzyme is available in the NCBI GenBank. The current understanding of nitrogenase diversity has been largely based on phylogenetic analyses of nitrogenase structural genes i.e. *nif*H and *nif*D genes [12, 13] and regulatory gene *nif*A [14]. However, limited work has been carried out on the annotation or phylogenetic analysis of *nif*H gene sequences and other *nif* genes or Nif proteins using computational and bioinformatics tools [15]. Thus, use of biological data mining and application of prediction tools for the computational identification of *nif* genes would further accelerate the research in the area of biological nitrogen fixation [16]. Moreover, the computational tools may also be useful to identify and categorize novel potential diazotrophs indicating evolutionary relationships [17].

2. RELATED WORK

Various *nif* genes involved in nitrogen fixation have been identified recently in different nitrogen-fixing bacteria [18, 19]. The nitrogenase enzyme consists of main structural gene *nif*H, which code for iron (Fe) protein (dinitrogenase reductase), whereas *nif*D and *nif*K genes code for iron-molybdenum (FeMo) protein [20, 21]. The protein encoded by *nif*H gene i.e., NifH protein, plays critical role in transfer of electrons during nitrogen fixation process. Therefore, *nif*H gene has been used as a molecular marker for identification of highly diversified nitrogen-fixing microbes from diverse environments [22]. Various *nif* genes involved in synthesis of nitrogenase enzyme in *Azospirillum brasilense* is represented (Fig. 1).

Quinto et al. [23] reported the complete coding sequence of the nitrogenase reductase gene (*nif*H) present in three different regions of a *Rhizobium phaseoli* symbiotic

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 02, February-2021

plasmid. The nucleotide sequences of the three *nif*H genes were found to be identical. Surveys of *nif*H diversity have been conducted in a wide range of environments including various marine [24] and terrestrial sites [25, 26]. The diversity of nitrogen-fixing organisms varies dramatically across habitats with different habitats selecting for different groups of nitrogen-fixing organisms [12, 27]. Enkh-Amgalan et al. [28] detected a major *nif* genes cluster in the strictly anaerobic, Gram-positive phototrophic bacterium *Heliobacterium chlorum*, consisting of 11 genes including *nif*H, *nif*D, *nif*K, *nif*E, *nif*N, *nif*B and *nif*V. The phylogenetic position of *Hbt. chlorum* nitrogenase reflected an evolutionary stage of a divergence of the two nitrogenase groups, with group I consisting of the aerobic diazotrophs and group II consisting of strictly anaerobic prokaryotes.

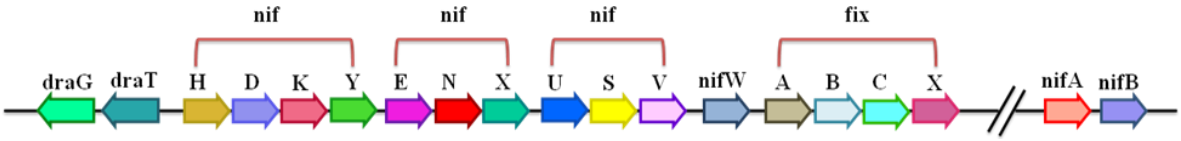

Fig. 1. Diagrammatic presentation of the *nif* genes of *Azospirillum brasilense. nif*H gene is located upstream of *nif*D gene on the nitrogenase encoding DNA

In similar phylogenetic studies, Choo et al. [29] used the complete DNA sequences of the three *nif*H genes in analysis of NifH phylogeny and demonstrated clustering of *Paenibacillus azotofixans* NifH1 and NifH2 within the *Cyanobacteriaceae* grouping. The NifH protein from *Trichodesmium* sp. strain IMS101 (marine cyanobacterium) showed the highest 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 *Methanothermococcus thermolithotrophicus* and *Methanothermobacter thermoautotrophicus*. Thus, phylogenetic analysis demonstrated that NifH1 and NifH2 form a monophyletic group among cyanobacterial NifH proteins, whereas NifH3 formed cluster among NifH proteins of the highly divergent methanogenic archaea. Frank [21] presented a novel approach to classify NifH protein sequences into well-defined phylogenetic clusters that provide a common platform for cross-ecosystem comparative analysis.

3. MATERIALS AND METHODS

Data mining of 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 [11, 30]. Phylogeny and phylogenetic trees give a picture of the evolutionary history among species, individuals or genes [31]. The need to query biological data using sets of evolutionarily related taxa has spawned the need to create databases than can serve as repositories of phylogenetic trees.

3.1 Retrieval of nifH gene sequences in different nitrogen-fixing bacteria

Basic Local Alignment Search Tool (BLAST) was used for searching of GenBank and other sequence databases for sequence similarity and homology among different nitrogen-fixing bacteria and nodule-forming rhizobia [32]. The search tool FASTA works on heuristic method of database searching and it uses a “hashing” submission of a query sequence and performed sequence for pairwise comparison of the query sequence with all individual sequences available in that database. In the present study, *nif*H gene nucleotide sequences from 40 nitrogen-fixing and nodule-forming bacterial strains were retrieved from NCBI GenBank. GenBank were accessed through the NCBI Entrez retrieval system and the NCBI homepage was used as the search point [33. BLAST was used for searching of GenBank and other sequence databases for sequence similarity and homology among different nitrogen-fixing and nodule-forming rhizobia. Phylogenetic analysis was carried out by using *nif*H gene sequences of *Azospirillum brasilense.*

3.2 Phylogenetic analysis of nifH gene sequences among different nitrogen-fixing bacteria

Phylogenetic analysis provides a visual means of representation for a group of sequences or species and indicates their time series of origin. Phylogeny and phylogenetic trees give a picture of the evolutionary history among species, individuals or genes [31]. If reliable phylogenies are produced, they will shed light on the sequence of evolutionary events that generated the present day diversity of genes and species, and help us to understand the mechanisms of evolution as well as the history of organisms. The phylogenetic study of nitrogen-fixing bacteria was carried out using *nif*H gene nucleotide sequences of *Azospirillum brasilense*. Datasets for nucleotide sequences, retrieved from NCBI GenBank, were created for *nif*H gene among different nitrogen-fixing and nodule-forming rhizobia. The filtered nucleotides sequences were aligned and the conserved region as well as 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 construction of phylogenetic trees [34]. Values above nodes represented bootstrap values.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 02, February-2021

Consensus phylogenetic trees were constructed for all sequences by by character based methods using Maximum Likelihood (ML) method [35]. 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 [36]. Phylogenetic tree is a two dimensional representation of relatedness among various biological species. It is a line drawing that provides a visual means of representation for a group of sequences or species and indicates their time series of origin. The phylogenetic tree is represented in three forms: Phylogram, Dendrogram, Cladogram. Consensus trees were constructed for all sequences using with and without bootstrapped method and the number of replications (iterations) used to construct the phylogenetic tree were taken as 1000 in MEGA (Fig. 2). 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. Phylogenetic relationships of genes or organisms usually are presented in a tree-like form with a root, which is called a rooted tree. Generated trees were viewed using TREE VIEW and best fit tree was selected out of all trees.

4. RESULTS

The nitrogen-fixing bacteria *Azospirillum* possess the ability to colonize the inner tissues of plants inter-cellularly. The understanding and optimizing these N_2-fixing plant-bacteria associations have promising prospective for sustainable agriculture. Their application as biofertilizer for cereals will reduce the energy and pollution cost of the industrial reduction of N_2 from fertilizer industry. Finally, 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 [22, 37, 38].

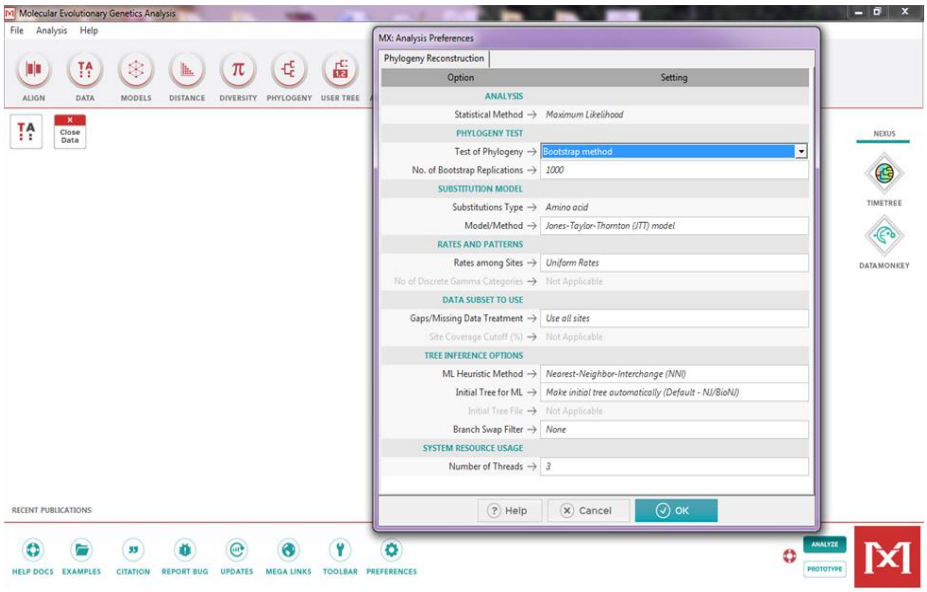

Fig. 2. Construction of phylogenetic trees of nucleotides with bootstrap method. The number of replications (iterations) used to construct the Phylogenetic Tree is 1000. The Statistical method used is Maximum Likelihood.

4.1 Sequence retrieval of nucleotide sequences of *nif*H gene sequences in nitrogen-fixing bacteria

In the present studies, 40 *nif*H gene sequences from different nitrogen-fixing bacterial strains and nodule-forming *Rhizobium* strains were retrieved from NCBI GenBank Database (Fig. 3). The *nif*H gene is one of the most heavily sequenced functional genes and >90% of *nif*H sequences in the database have been deposited since 2005. Analysis of database content showed that there are 375 full-length *nif*H genes in the database including 245 from sequenced genomes. In this study, it was observed that nucleotide sequences of *nif*H gene varied in different nitrogen-fixing bacteria. For example, *Azospirillum brasilense* (endophytic bacterium) contained 1826 nucleotide bases, *Bradyrhizobium japonicum* (soybean nodulating bacteria) possessed only 1280 nucleotides, whereas *Klebsiella pneumoniae* (free-living bacteria) was found to contain 1807 nucleotide bases.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 02, February-2021

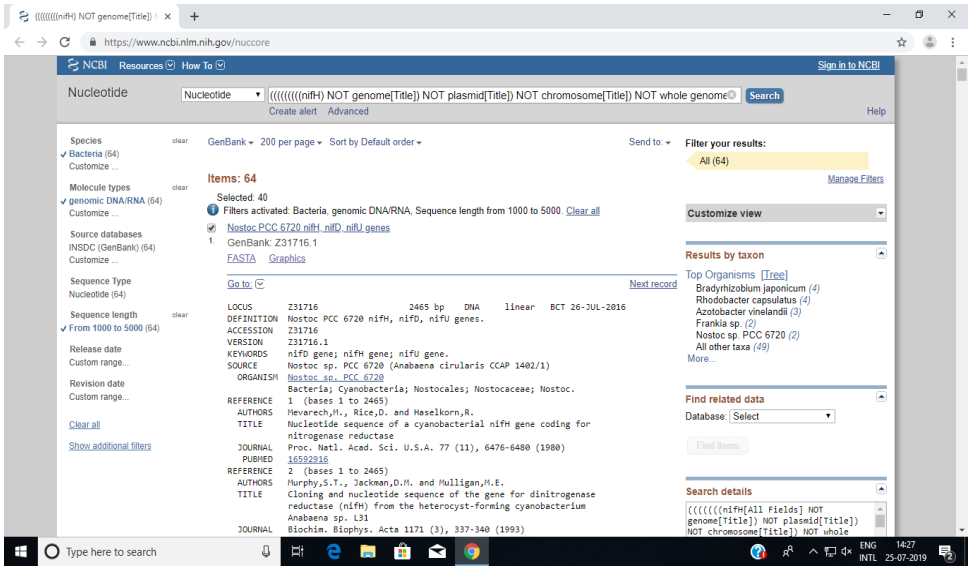

Fig. 3. Sequence retrieval of *nif*H gene using query in NCBI

DNA contains various genes, which code for different proteins and enzymes to perform various metabolic functions. The nucleotide sequences of *nif*H gene in nitrogen-fixing organism *Azospirillum brasilense* in FASTA format are provided below. Nucleotide sequence G represents guanine, A for adenine, C for cytosine and T represents thymine base. Three nucleotide bases constitute a codon during transcription and finally get translated to make specific amino acid depending upon the nucleotide bases present in the DNA sense strand.

>*Azospirillum brasilense* (1)

AACGCCTTCGTCCGAGCCCGCCGACCCGCAAGGCGATGAACGATTGCCTGTCGCGCTGCGGCTTTCCGACACGCTGT
TCGATCTTGTCGGAAAGGCGGCCAACGCGCGACCTCCGACGAGCTGCTGTTCGAGGCGCTGCTGCGCCAGCTCCGTT
CCCTGCACCAGTTCTGCACGGATCTCGACCATGCCGGGCGTCCGCGGGTGCTCGACACGCTGGAACGCCGTCTGGCC
GAACGCAGCCGCGCGCTGGCCGGCATCGCCGGGGAGCGCCTCGTCGGCTCGTCCTGCTGGCCCGCCCCGCCGACCCG
GCCAAGGCGCGCCGCCTGCTCGCCCGCGGCCAGTCGCTGGCCCAGCTTCTCTATGACATGGGCCAGGACGGCGACGA
GTTGGAGGCGTTGGCGCTGCGCCTCGTCGTCGCCCGCGACGCGCTGAACCACGCCGCCGCCTGACGCTGGCTCTGAC
GCTGGCGGACCGACATTCGTCCATCCCTTCCTTTTCATTGCGAGAGACCCGACAAAATGTCGGGCTTTGTCGGGTAT
GTCACAGGCCCGACAAAGCGGACCGCGCTCGTCCGCAATCAATTTTCTTGTTCAATTTCAATGATTTAAAAATTTTT
CGCGAACTGGCACGGGGGATGCAGAGAAGGGGTCGAAGCGGCCGCTGGCAGGCCGCCCCGGAATTGAAGACACCCTG
TAGACCCAAGCAAAGGAGTAACCTCCCATGTCTTTGCGCCAGATTGCGTTCTACGGTAAGGGCGGTATCGGCAAGTC
CACCACCTCCCAGAACACCCTGGCCGCGCTGGTCGAGCTGGATCAGAAGATCCTGATCGTCGGCTGCGATCCGAAGG
CCGACTCGACCCGCCTGATCCTGCACGCCAAGGCGCAGGACACCGTGCTGCACCTCGCCGCCGAAGCCGGCTCGGTC
GAGGATCTGGAGCTCGAGGACGTTCTCAAGATCGGCTACAAGGGCATCAAGTGCGTCGAGTCCGGCGGTCCGGAGCC
GGGGGTCGGCTGCGCCGGCCGCGGCGTGATCACCTCGATCAACTTCCTGGAAGAGAACGGCGCCTACGACGACGTGG
ACTACGTCTCCTACGACGTGCTGGGCGACGTGGTGTGCGGCGGTTTCGCCATGCCCATCCGCGAGAACAAGGCCCAG
GAAATCTACATCGTCATGTCCGGTGAGATGATGGCGCTCTACGCCGCCAACAACATCGCCAAGGGCATTCTGAAGTA
CGCCCACAGCGGCGGCGTGCGCCTCGGCGGCCTGATCTGCAACGAGCGCCAGACCGACAAGGAAATCGACCTCGCCT
CCGCCCTGGCCGCCCGCCTCGGCACCCAGCTCATCCACTTCGTGCCGCGCGACAACATCGTGCAGCACGCCGAGCTG
CGCCGCATGACGGTGATCGAGTACGCGCCGGACAGCCAGCAGGCCCAGGAATACCGCCAGCTCGCCAACAAGGTCCA
CGCGAACAAGGGCAAGGGCACCATCCCGACCCCGATCACGATGGAAGAGCTGGAAGAGATGCTGATGGACTTCGGCA
TCATGAAGTCGGAGGAGCAGCAGCTCGCCGAGCTCCAGGCCAAGGAAGCCGCCAAGGCCTGATAACTGGGTCCCCCA
CGGGGCGCAGTTCTGAACTGGCCCCCTCCCCAGCCCACCCCGCCCCTGTGCGCGGGGTGGGAGAGAGGGGGCTCGGT
CCCGCGCTGCAGTAGCGCGGATAATGAGTGCAGCAGGAGACTGGCACCATGAGCCTGTCCGTGAACGAAGGCGTCGA
CGTCAAGGGTCTCGTCGACAAGGTTCTCGAAGCGTATCCCGAGAAGTCGCGCAGG

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 02, February-2021

4.2. Phylogenetic trees construction based on the sequence similarity index of *nif*H gene sequences among different nitrogen-fixing bacteria

Since all biological entities have come about through the evolutionary process, the patterns, functions and processes that they possess are best analyzed in terms of their phylogenetic histories [35]. These changes along a branch affect the biology of all descendant species, thereby leaving phylogenetic patterns everywhere. The *nif*H gene is a widely used molecular proxy for studying nitrogen fixation and phylogenetic classification of *nif*H gene sequences is an essential step in diazotroph community analysis. In this study, the level of similarity was searched among the nucleotide sequences of nitrogen-fixing bacteria. Alignment of all retrieved sequences were done using CLUSTALW program. Conserved region and region of dissimilarity were identified using multiple sequence alignment. Phylogenetic trees were generated by Maximum Likelihood method.

The *nif*H gene sequences of *Azospirillum brasilense* showed close similarity with the *Rhodobacter capsulatus* and *Rhodospirillum rubrum* (Fig. 4), suggesting that *nif*H gene sequences lies within the Rhodobacterial clade. Further, the sequences also showed relatedness to nodule-forming bacteria i.e., *Rhizobium leguminosarum* (pea nodulating bacteria), *Sinorhizobium meliloti* (alfalafa nodulating), *Bradyrhizobium japonicum* (soyben nodulating) and *Azorhizobium caulinodans* (which make nodules on root as well as stem of *Sesbania*). On the another branch, *nif*H gene sequences of *Azospirillum brasilense* showed relatedness to *Frankia* sp. (actinrhizal nodules) and showed similarity with free-living nitrogen-fixing bacteria *Klebsiella* sp. and *Azotobacter* sp. Further observations and the branching order of the nifH phylogeny showed that nitrogen-fixing bacteria isolated from diverse environments were distantly related. For example, *Cyanobacterium* sp., *Trichodesmium erythraeum*, *Mastigocladus laminosus*, *Paenibacillus sabinae* and *Nostoc commune* were placed on different branch showing large variations in their nucleotide sequences.

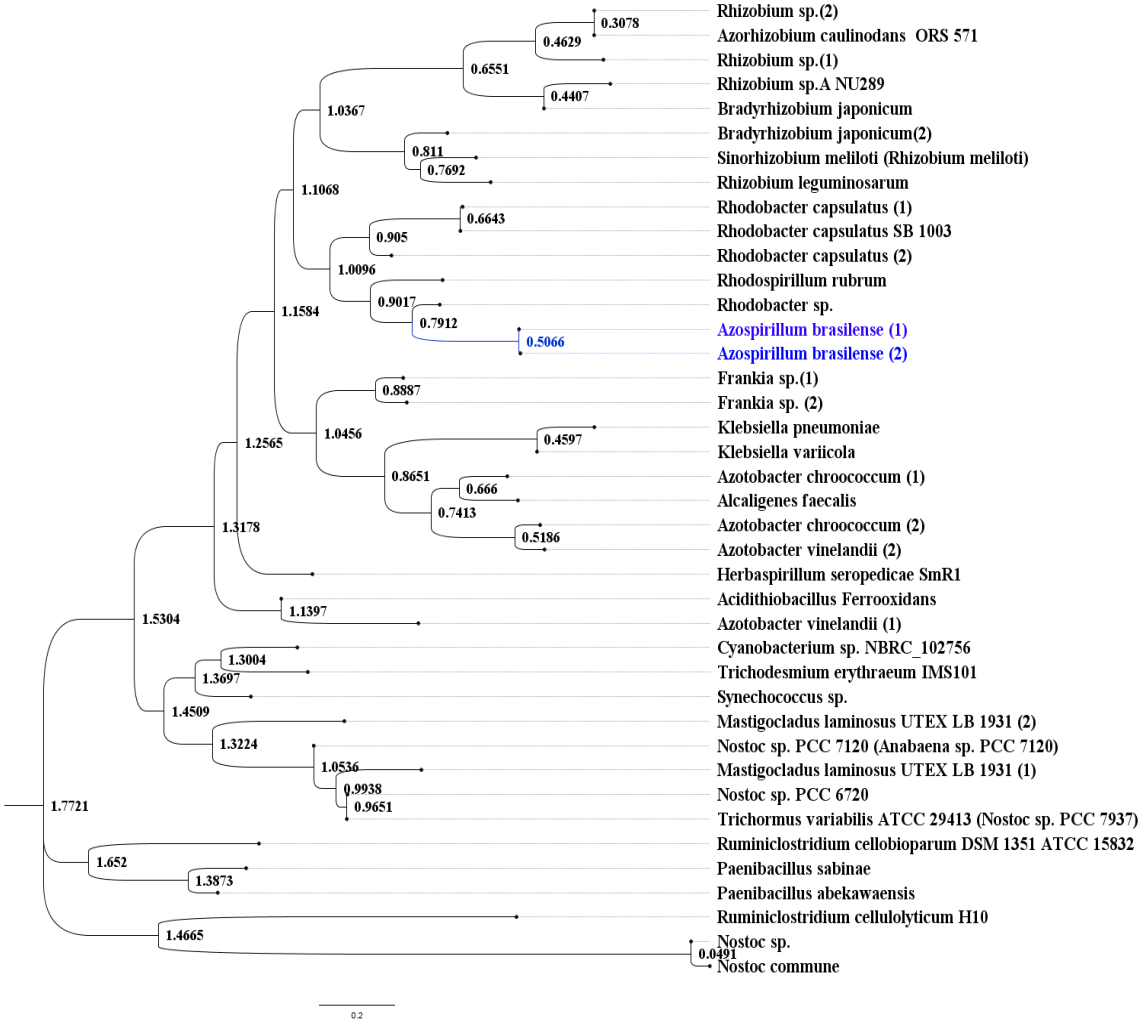

Fig. 4. Phylogenetic tree of *nif*H gene by Maximum Likelihood method using MEGA-X

In the Maximum Likelihood method, the bootstrapping values indicate how many times out of 1000 the same branch was observed when repeating the phylogenetic reconstruction on a re-sampled set of your data. The node ages were

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 02, February-2021

generally taken in decreasing order i.e. 10, 9, 8, 7, 6, 5, 4, 3, 2, 1 that depicted that lower the value of node age, the organism on that taxa was evolutionary evolved early in the history. Using bootstrapping in Maximum Likelihood method, the nucleotide sequence database similarity of the *nif*H gene sequences were compared with other nitrogen-fixing bacteria. It was observed that the *nif*H gene sequences of *Azospirillum brasilense* showed close similarity with the *Rhodospirillum rubrum* and *Rhodobacter capsulatus* (Fig. 5), suggesting that *nif*H gene sequences lies within the Rhodobacterial clade. Further, the sequences also showed relatedness to nodule-forming bacteria i.e., *Bradyrhizobium japonicum* (soybean nodulating), *Rhizobium leguminosarum* (pea nodulating bacteria), *Sinorhizobium meliloti* (alfalafa nodulating), and *Azorhizobium caulinodans* (which make nodules on root as well as stem of *Sesbania*). *nif*H gene sequences were placed on another branch in *Trichomonas variabilis*, *Nostoc* sp., *Trichodesmium erythraeum* and *Herbaspirillum seropodicae.* Large variations in the nucleotide sequences were observed in *Azotobacter vinelandii*, *Azotobacter chroococcum*, *Paenibacillus sabinae* and *Nostoc commune* and these bacteria were placed on distant branch.

5. DISCUSSION

The creation of centralized and vetted *nif*H databases of 16S ribosomal RNA (rRNA) gene sequences is the urgent need for their utility in ecology of nitrogen-fixing microorganisms [39, 40]. Therefore, *nif*H gene is most widely used molecular marker to study the ecology and evolution of nitrogen-fixing prokaryotes in nature [2]. The distribution pattern of nitrogen-fixing bacteria suggests that nitrogen-fixing ability is evolutionary ancient and mainly transmitted vertically with the widespread loss of function [41]. The recent rapid expansion of microbial genome sequences has revealed the presence of the genes encoding homologous proteins to known nitrogenases, even in prokaryotic species that had not previously been recognized as diazotrophs [17].

In earlier studies, five main phylogenetic clusters were reported based on the homologs of the *nif*H gene [42]. Aerobic and facultatively anaerobic organisms constitute 'Cluster I of *nif*H genes' that belong to phyla including Proteobacteria, Cyanobacteria, Firmicutes and Actinobacteria. Cluster II contains *anf*H, alternative nitrogenases that are paralogs of *nif*H and use Fe–Fe cofactor in place of the Fe–Mo cofactor used by *nif*H [43]. *nif*H genes belonging to Cluster III are almost exclusively found in obligate anaerobes including methanogenic Archaea, Treponema, *Clostridium* and sulfate-reducing and sulfur-reducing species of Delta-proteobacteria. The alternative nitrogenases containing V–Fe (encoded by *vnf*H gene) appear to be found only in the genomes of organisms that also contain *nif* genes [43]. Clusters IV and V contain paralogous genes that do not participate in nitrogen fixation [2, 44].

Phylogeny and phylogeography of functional genes was performed among seven terrestrial microorganisms living in extreme environments [45]. Seven metagenomic libraries were produced from fracture water samples gathered from five South African mines. Metagenomic data analysis for evolutionary study of eight ubiquitous functional genes i.e. *nar*V, NPD, PAPS reductase, *nif*H, *nif*D, *nif*K, *nif*E and *nif*N genes, showed that functional genes were taxonomically and phylogenetically diverse and distinct from one another. But, dissimilarity between samples did not connect strongly with geographical and environmental parameters. Frank et al. [22] reported that phylogenetic classification of *nif*H gene sequences is an essential step in diazotroph community analysis that requires a fast automated solution due to increasing size of environmental sequence libraries. Thus, rapid and automated phylogenetic cluster assignment circumvents extensive phylogenetic analysis of *nif*H sequences and it saves substantial time and resources in nitrogen fixation studies.

The phylogenies of the *nod*Y/K and *nif*H genes of 45 *Bradyrhizobium* strains isolated from different legumes were compared on the basis of their 16S rRNA gene phylogeny and genetic diversity [46]. Strains were distributed into two superclades - *B. japonicum* and *B. elkanii* - with several strains being very similar within each clade in the 16S rRNA tree. Gaby et al. [38] developed a computational pipeline which infers taxonomy and optionally filters out paralog sequences and reported that *nif*H gene may be used as the most widely established molecular marker for the study of nitrogen-fixing prokaryotes in nature. Recently, Sindhu et al. [14] performed phylogenetic analysis of diazotrophic bacteria using *nif*A gene nucleotide sequences of *Azorhizobium caulinodans* (*Sesbania* nodulating bacteria) and reported that *A. caulinodans* showed similarity with the *Rhodoblastus sphagnicola*, *Rhodoblastus acidophilus* and *Rhodopila globiformis.* In addition, other nodule-forming bacteria including *Mesorhizobium ciceri*, *Mesorhizobium mediterraneum* and *Bradyrhizobium japonicum* alongwith associative symbionts *Azospirillum lipoferum* and *Azospirillum brasilense* were found closely related. On the other hand, nodule-forming bacteria such as *Rhizobium leguminosarum*, *Rhizobium etli* and *Sinorhizobium meliloti*, along with free-living nitrogen-fixing bacteria *Klebsiella pneumoniae* and *Azotobacter vinelandii* were found distantly related.

6. CONCLUSION

In intensive agriculture, high yielding crop varieties are grown by farmers to feed the ever-increasing human population. Nitrogen is the most limiting nutrient for growth of these cereal and leguminous plants [47], which is provided in the soil mostly through application of inorganic nitrogenous fertilizers. However, excessive use of these chemical fertilizers has deteriorated soil health and fertility in different agro-ecosystems and polluted the environment leading to various health hazards. Therefore, use of nitrogen fixing bacteria as biofertilizer has emerged as an ecofriendly and cost-effective technology for achieving sustainable restoration of soil fertility [48]. Considering the importance and application of nitrogen fixing bacteria in agricultural fields, the present study was undertaken for phylogenetic analysis of *nif*H gene sequences in *Azospirillum brasilense*, which fixes nitrogen in association or within plant tissues.

In the present study, *nif*H gene sequences from 40 different nitrogen-fixing bacterial species were retrieved using

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 02, February-2021

NCBI GenBank. Close similarity of *Azospirillum brasilense* *nif*H gene sequences was observed with the *Rhodobacter capsulatus* and *Rhodospirillum rubrum*, where as nitrogen-fixing bacteria isolated from diverse environments were distanly related. For example, *Cyanobacterium* sp., *Trichodesmium erythraeum*, *Mastigocladus laminosus*, *Paenibacillus sabinae* and *Nostoc commune* were placed on different branch showing large variations in their nucleotide sequences. Using bootstrapping, *nif*H gene sequences of *Azospirillum brasilense* again showed close similarity with the *Rhodospirillum rubrum* and *Rhodobacter capsulatus*. Further, the sequences also showed relatedness to nodule-forming bacteria i.e., *Bradyrhizobium japonicum*, *Rhizobium leguminosarum*, *Sinorhizobium meliloti* and *Azorhizobium caulinodans*. Large variations in the nucleotide sequences were observed in *Azotobacter vinelandii*, *Azotobacter chroococcum*, *Paenibacillus sabinae* and *Nostoc commune* and these bacteria were placed on distant branch. The present studies would be very helpful to identify, verify and classify various types of nitrogen-fixing bacteria. Finally, this study has shown the correlation or phylogenetic relatedness among the *nif*H gene sequences obtained from different taxa.

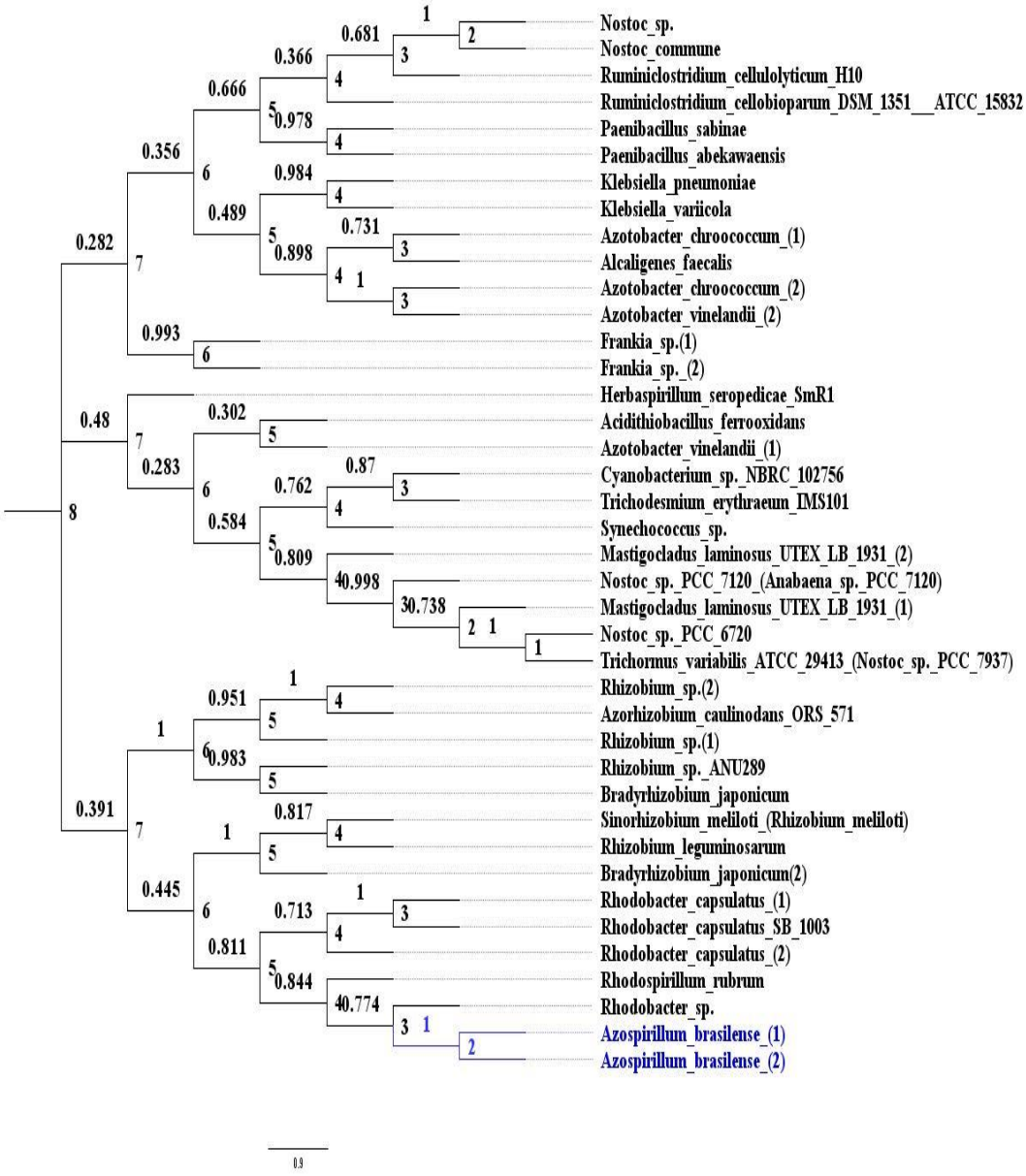

Fig. 5. Phylogenetic tree of *nif*H gene using bootstrap method (Number of iterations = 1000) by Maximum Likelihood method in MEGA-X

REFERENCES

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

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

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

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

[4] P.M. Vitousek, J.D. Aber, R.W. Howarth, G.E. Likens, P.A. Matson et al., "Human alteration of global nitrogen cycle: sources and consequences". Ecology and 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 Biogeochemical Cycles, Vol. **13**, pp. **623-645**, **1999**.

[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] 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**.

[8] Y. Okon, C. Labandera-Gonzales, 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**.

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

[10] J.H. Kim, "Next-generation sequencing technology and personal genome data analysis". In: Genome Data Analysis. Learning Materials in Biosciences. Springer, Singapore, pp. **17-31**, 2019.

[11] P.M. Selzer, R.J. Marhöfer, O. Koch, "The biological foundations of Bioinformatics". In: Applied Bioinformatics. Springer, Cham. pp. **1-12**, **2018**.

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

[13] B.J. Henson, L.E. Watson, S.R. Barnum, "The evolutionary history of nitrogen fixation, as assessed by *nif*D". Journal of Molecular Evolution Vol. **58**, pp. **390–399**, **2004**.

[14] D. Sindhu, S. Sindhu, S.K. Yadav, "Data mining and phylogenetic analysis of nifA gene sequences using nodule-forming *Azorhizobium* strain", International Journal of Emerging Trends and Techology in Computer Sciences. Vol. **10**, pp. **1-10**, **2021**.

[15] 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 Biomolecule Structure Dyn, Vol. **26**, pp. **321-328**, **2008**.

[16] P.K. Meher, T.K. Sahu, J. Mohanty, S. Gahoi, S. Purru, M. Grover, A.R. Rao, "*nif*Pred: 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**.

[17] 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**.

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

[19] Q. Li, X. Liu, H. Zhang, S. Chen, "Evolution and functional analysis of *orf*1 within *nif* gene cluster from *Paenibacillus graminis* RSA19". International Journal of Molecular Sciences, Vol. **20**, pp. **1145**, **2019**.

[20] 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**.

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

[22] I. Frank, K.A. Turk-Kubo, J.P. Zehr, "Rapid annotation of *nif*H gene sequences using classification and regression trees facilitates environmental functional gene analysis". Environmental Microbiology Reports, Vol. **8**, Issue **5**, pp. **905–916**, **2016**.

[23] C. Quinto, H. de la Vega, M. Flores, J. Leemans, M.A. Cevallos, M.A. Pardo, et al., "Nitrogenase reductase: functional multigene family in *Rhizobium phaseoli*". Proceedings of Environmental Academy of Sciences, USA, Vol. 82, pp. **1170-1174**, **1985**.

[24] H. Farnelid, A.F. Andersson, S. Bertilsson, W. Abu Al-Soud, L.H. Hansen, S. Sorensen, et al., "Nitrogenase genes amplicons from global marine surface waters are dominated by genes of non-cyanobacteria". Plos One Vol. **6**: e19223, **2011**.

[25] 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**.

[26] J.A. Izquierdo, K.Nusslein, "Distribution of extensive *nif*H gene diversity across physical soil microconditions". Microbial Ecology, Vol. **51**, pp. **441–452**, **2006**.

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

[28] J. Enkh-Amgalan, H. Kawasaki, T.Seki, "Molecular evolution of *nif* gene cluster carrying $nifI_1$ and $nifI_2$ genes in Gram-positive phototrophic bacterium *Heliobacterium chlorum*". International Journal of Systemic and Evolutionary Microbiology, Vol. **56**, pp. **65-74**, **2006**.

[29] Q.-C. Choo, M.-R. Samian, N. Najimudin, "Phylogeny and characterization of three *nif*H-homologous genes from *Paenibacillus azotofixans*". Applied and Environmental Microbiology, Vol. **69**, Issue **6**, pp. **3658–3662**, **2003**.

[30] 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**.

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

[32] 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**.

[33] E.W. Sayers, T. Barrett, D.A. Benson, S.H. Bryant, K. Canese et al., "Database resources of National Center for Biotechnology Information", Nucleic Acids Research, Vol. **38**, pp. **D5–D16**, **2010**.

[34] 34 S. Kumar, K. Tamura, M. Nei, "MEGA: Molecular Evolutionary Genetics Analysis. Ver. 1.0". The Pennsylvania State University, University Park, PA, **1993**.

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

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

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

[38] 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**, pp. **e01512-17**, **2018**.

[39] T.Z. DeSantis, P. Hugenholtz, N. Larsen et al., "Greengenes, chimera-checked 16S rRNA gene database and workbench compatible with ARB". Applied and Environmental Microbiology, Vol. **72**, pp. **5069–5072**, **2006**.

[40] J.R. Cole, Q. Wang, E. Cardenas, et al., "Ribosomal database project: improved alignments and new tools for rRNA analysis". Nucleic Acid Research, Vol. **37**, pp. **D141–D145**, **2009**.

[41] 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**.

[42] Y.T. Chien, S.H. Zinder, "Cloning, DNA-sequencing and characterization of *nif*D-homologous gene from archaeon *Methanosarcina barkeri*-227 which resembles *nif*D from eubacterium *Clostridium pasteurianum*". Journal of Bacteriology, Vol. **176**, pp. **6590–6598**, **1994**.

[43] B. Masepohl, K. Schneider, T. Drepper et al., "Alternative nitrogenases". In: Nitrogen Fixation at Millennium, G.J. Leigh (ed.). Elsevier Science B.V., Amsterdam, Netherlands, pp. **191–222**, **2002**.

[44] Y. Fujita, Y. Takahashi, M. Chuganji et al., "*nif*H-like (*frx*C) gene is involved in biosynthesis of chlorophyll in filamentous cyanobacterium *Plectonema boryanum*". Plant Cell Physiology, Vol. **33**, pp. **81–92**, **1992**.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 02, February-2021

[45] M.C.Y. Lau, C. Cameron, C. Magnabosco, C.T. Brown 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**, Article 531, pp. **1-17**, **2014**.

[46] J.R. Delamuta, P. Menna, R.A. Ribeiro, M. Hungria, "Phylogenies of symbiotic genes of *Bradyrhizobium* symbionts of legumes of economic and environmental importance in Brazil support definition of new symbiovars *pachyrhizi* and *sojae*". Systemic Applied Microbiology, Vol. **40**, Issue **5**, pp. **254-265**, **2017**.

[47] G.A. Divito, V.O. Sadras, "How do phosphorus, potassium and sulphur affect plant growth and biological nitrogen fixation in crop and pasture legumes"? meta-analysis. Field Crop Research, Vol. **156**, pp. **161–171**, **2014**.

[48] 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**.

AUTHORS PROFILE

Dr. 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 24 research articles in national and international journals. He has participated in 17 national / international seminars and conferences. He has successfully completed his Ph. D. degree from Shri JJT University, Jhunjhunu, Rajasthan. He is having five years of teaching experience at the post of Assistant Professor in CRM Jat College, Hisar.

Neeru Singh Redhu received her Bachelor of Science degree from Kurukshetra University, Kurukshetra in 2007. She successfully completed her Masters' of Science degree in Bioinformatics from CCS Haryana Agricultural University, Hisar in 2009. Her research interests mainly include Bioinformatics, sequence analysis, phylogenetic studies, protein structure prediction and protein modeling among others. She has published 7 research articles in national and international journals. She has participated in 10 national/international seminars and conferences. She joined at the post of Assistant professor in 2011 at CCS Haryana Agricultural University, Hisar and has been teaching Bioinformatics courses since then. Currently, she is pursuing her Ph. D. degree in Molecular Biology and Biotechnology from the CCS Haryana Agricultural University, Hisar.

Dr. Divya Sindhu received the B. Tech. degree in Computer Science and Engineering from Kurukshetra University, Kuruksheta in 2012. She has 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 26 research articles in national and international journals. She has participated in 18 national / international seminars and conferences. She has successfully completed her Ph. D. degree from Shri JJT University, Jhunjhunu, Rajasthan. She is having five years of teaching experience at the post of Assistant Professor in 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.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)