International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)
Web Site: www.ijettcs.org Email: editor@ijettcs.org, editorijettcs@gmail.com
Volume 10, Issue 1, January - February 2021 ISSN 2278-6856

Data mining and phylogenetic analysis of nifA gene sequences using nodule-forming Azorhizobium strain

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

[1*]Ph. D. Scholar, Department of Computer Science and Engineering
Shri Jagdishprasad Jhabarmal Tibrewala University, Jhunjhunu - 333001, Rajasthan, India

[2]Ph. D. Scholar, Department of Computer Science and Engineering
Shri Jagdishprasad Jhabarmal Tibrewala University, Jhunjhunu - 333001, Rajasthan, India

[3]Professor, Department of Computer Science and Engineering
Shri Jagdishprasad Jhabarmal Tibrewala University, Jhunjhunu - 333001, Rajasthan, India

Abstract: In the post-genomic era involving novel sequencing technologies, different nitrogen fixation (*nif*) genes encoding nitrogenase enzyme proteins have been identified through comparative genomic and mutational analysis along with transcriptional profiling. In earlier studies, *nif*H gene sequences have been used as a molecular marker for the phylogenetic distribution of diazotrophs. In the present study, *nif*A gene sequences from *Azorhizobium caulinodans* (*Sesbania* nodulating bacteria) and 24 different nitrogen-fixing bacterial species were retrieved from NCBI GenBank and phylogenetic analysis was performed using Maximum likelihood method. *nif*A gene nucleotide sequences of *Azorhizobium caulinodans* showed similarity with the *Rhodoblastus sphagnicola, Rhodoblastus acidophilus* and *Rhodopila globiformis.* Other nodule-forming bacteria *Mesorhizobium mediterraneum, Mesorhizobium ciceri* and *Bradyrhizobium japonicum* were found closely related, whereas *Rhizobium leguminosarum, Rhizobium etli* and *Sinorhizobium meliloti* were found distantly related. The associative symbionts i.e., *Azospirillum lipoferum* and *Azospirillum brasilense* were also found closely related, whereas free-living nitrogen-fixing bacteria *Klebsiella pneumoniae* and *Azotobacter vinelandii* were distantly related. Using boot strap method, *Rhodoblastus sphagnicola, Rhodoblastus acidophilus* and *Rhodopseudomonas palustris, Azospirillum lipoferum* and *Azospirillum brasilense* and *Bradyrhizobium japonicum* were found closely related with *nif*A sequences of *Azorhizobium caulinodans*. On the other hand, *Sinorhizobium meliloti, Rhizobium leguminosarum, Klebsiella pneumoniae* and *Azotobacter vinelandii* as well as *Rhodobacter capsulatus* were found distantly related.

Keywords: Data mining, Phylogenetic analysis, Nitrogen fixation, *nif*A gene, Nucleotide sequences, *Azorhizobium caulinodans*

1. INTRODUCTION

Currently, large amounts of biological data have been generated by the scientific community due to rapid developments in genomics and proteomics. To draw conclusions from these biological data, sophisticated computational analyses are required using data mining (DM) approaches. Thus, interdisciplinary science of Bioinformatics or Computational Biology has evolved tremendously in recent years [1], [2]. The prediction tools have been developed for the computational identification of nitrogen fixation (*nif*) genes and categorization of potential diazotrophs using high throughput sequence data in the area of biological nitrogen fixation [3]. Recently, *in silico* analysis of genes and proteins has been receiving greater attention with particular emphasis to find suitable biomarkers for rapid identification of different pathogenic genera [4] and discovery of potent microbial enzymes useful for several agricultural and animal feed industries [5], [6].

In sustainable agriculture, biological nitrogen fixation (BNF) offers the alternative to addition of nitrogenous fertilizers in cropping systems, as it uses the capacity of certain nitrogen-fixing bacteria to fix atmospheric nitrogen into the plant usable ammonia using the enzyme nitrogenase [7]. Nitrogen fixation is widely but sporadically distributed among both eubacteria and methanogenic archaea [3], [8]. The computational tools are recently being used for the annotation and phylogenetic analysis of *nif*H gene or NifH protein sequences. Bacterial conversion of nitrogen (N_2) to ammonia (NH_3) is an energetically expensive process and sensitive to oxygen, ammonia and combined nitrogen (added as fertilizer in the soil). Thus, nitrogenase synthesis is switched off under aerobic and nitrogen-sufficient conditions. Moreover, species-specific environmental and metabolic conditions are essential for the manifestation of biochemical pathways [9]. Due to their inceptive role in nitrogen cycle, diazotrophs are present in virtually all ecosystems including aerobic soils (e.g., *Azotobacter* species), the ocean surface layer (*Trichodesmium*) and specialized nodules in legume

roots (*Rhizobium*). The need to query data using sets of evolutionarily related taxa has spawned the need to create databases for their use as repositories of phylogenetic trees. Phylogeny and phylogenetic trees give a picture of the evolutionary history among species, individuals or genes [10].

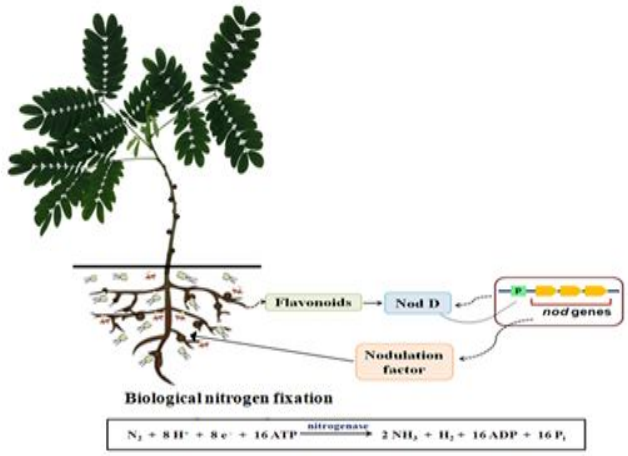

Figure 1 Stem nodules formed by Azorhizobium cuulinoduns strain ORS571 on Sesbania rostrata. The process of nodule formation and nitrogen fixation has also been represented diagrammatically

Considering the importance of nitrogenase enzyme in agriculture [11], [12], the genes involved in nitrogen fixation (nif and fix genes) have been identified in the free-living and symbiotic rhizobia [13], [14]. Further, phylogenetic analysis of nifA nucleotide sequences was carried out in Azorhizobium caulinodans, which make nodules on root as well as stem of Sesbania (Fig. 1). With its dual nodulation topology on both stems and roots of S. rostrata, Azorhizobium caulinodans offers a unique system for investigating the interaction with the host legume, which is phylogenetically separated from other rhizobia [15], [16].

The current understanding of nitrogenase diversity has been based largely on phylogenetic analyses of nifH and nifD, the nitrogenase structural genes [17], [18]. Recently, Raymond et al. [3] performed genomic analyses of nif genes encoding the core components of nitrogenase, including the NifH, NifD, NifK, NifE and NifN proteins. This grouping was largely consistent with the previous classification, in which the nitrogenase genes were divided into clusters I–IV [17]. nifA gene product acts as a positive activator of the transcription of nif operons, whereas nifL gene product is involved in the negative control in response to environmental and metabolic conditions for the manifestation of nitrogen fixation pathway [9, 19]. Till now, nifH gene sequences are being used as molecular marker for identification of nif genes and phylogenetic analysis among different rhizobia and in other nitrogen-fixing bacteria. To the best of our knowledge this is the first report in which nifA gene has been used for the computational prediction of the of nitrogen-fixation (nif) genes in diazotrophs. Thus, detailed nucleotide sequences of nifA genes using in silico methodologies and bioinformatics tools will facilitate phylogenetic and evolutionary studies of nitrogen-fixing microorganisms [20].

2. 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 [1]. Phylogeny and phylogenetic trees give a picture of the evolutionary history among species, individuals or genes [10]. 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.

2.1. Retrieval of nifA 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 [21]. Thus, BLAST can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families. To access GenBank and its related retrieval and analysis services, the NCBI homepage was used as the search point [22]. 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, 24 nifA gene sequences from different nitrogen-fixing and nodule-forming bacterial species were retrieved from NCBI GenBank (Fig. 2). GenBank were accessed through the NCBI Entrez retrieval system. BLAST was used for searching of GenBank and other sequence databases for sequence similarity and homology among different nodule-forming rhizobia. GenBank and its related retrieval and analysis services were accessed using the NCBI homepage. Phylogenetic analysis, with other nitrogen-fixing bacteria and nodule-forming rhizobial species, was done by taking nifA gene sequences of Azorhizobium caulinodans, which makes nodules on both root and stem of Sesbania crops.

2.2. Phylogenetic analysis of nifA gene sequences among different nitrogen-fixing bacteria using computational software

Phylogenetic analysis provides a visual means of representation for a group of sequences or species and indicates their time series of origin. In the present study, phylogenetic study of nitrogen-fixing bacteria was carried out using nifA gene nucleotide sequences of Azorhizobium caulinodans. Nucleotide sequences of nifA gene were searched for different nitrogen-fixing and nodule-forming rhizobia from NCBI GenBank databases. Datasets for both nucleotide sequences, retrieved from NCBI GenBank were

created for nifA gene. Partial sequences for nucleotides were removed from retrieved sequence datasets. 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 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. Values above nodes represented bootstrap values.

Consensus phylogenetic trees were constructed for all sequences by by character based methods using Maximum Likelihood (ML) method [23] 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 [24].

Based on the nucleotide sequence database similarity, the relatedness of different nucleotide sequences of nifA gene 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. The program has a graphical interface that allows users to modify various components of the tree such as rooting positions, node labels, tip labels and scale axes. 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.

3. RESULTS

NifA protein plays a major role in transcriptional activation of nif and fix genes [19]. MEGA software (i.e., MEGA-X) was used for conducting statistical analysis of molecular evolution and for constructing phylogenetic trees. The Statistical method used was Maximum Likelihood.

3.1. Sequence retrieval of nifA gene in nitrogen-fixing and nodule-forming bacteria

In this study, the nucleotide sequences of nifA gene from 24 nitrogen-fixing bacterial strains were retrieved from NCBI GenBank in FASTA format for computational analysis [22, 25]. NCBI BLAST was used for searching of GenBank and other sequence databases for sequence similarity and homology among different nitrogen-fixing and nodule-forming rhizobia (Fig. 3). The bacteria used for analysis of nifA gene nucleotide sequences had been isolated

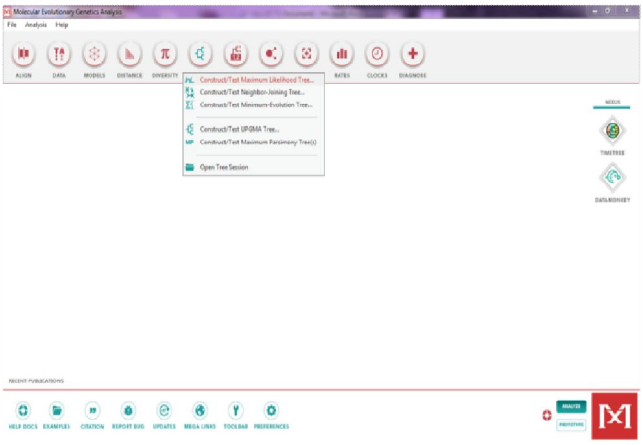

Figure 2 Different methods used to construct phylogentic Trees in MEGA-X

from differentat habitats and range from aerobic to anaerobic life style, free-living to nodule-forming symbiotic bacteria.

The nucleotide sequences of *nif*A gene in different nitrogen-fixing organisms were opened in FASTA format and is provided below. DNA contains various genes, which code for different proteins and enzymes to perform various metabolic functions. 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.

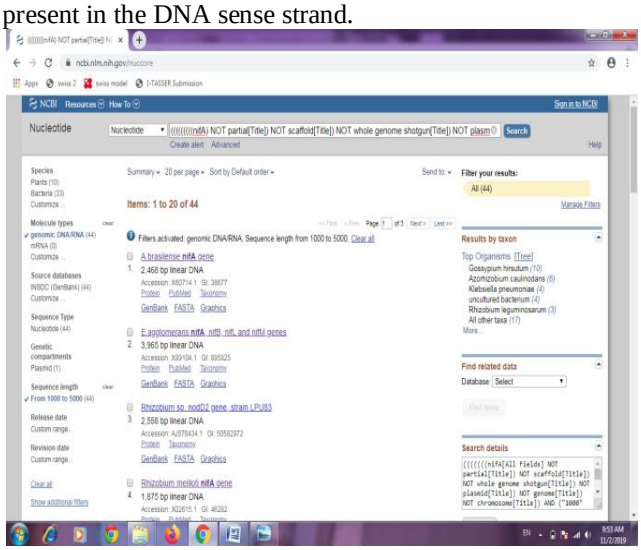

Figure 3 Sequence retrieval of *nif*A gene

>***Azorhizobium caulinodans*** **ORS 571**

GAACAGCAGCTCTACGGGCTGCTGGCTGCGGCTAC
ACCACGCGCAACGGCACGGTCGCGCAGGCACTGG
AGGCAGAGGCCGCCCCCCACCTTTCGCCGGCCGCG
CTCGACGCGGCGAAGGCGGCGGCCTCCATCATGG
CCATGAACAACGTCTACTATCGATTTACGCATCTG
GCCTCCAACAAGGCCTATGAGACCCTGCCGGCGA
AGCTCCGCATGAGCGTCATCGGCAATCCCGGCGTG

GACAAGGTGGACTTCGAACTCTGGTCGCTCGCCGT
CTCGGCCATGAACGGATGCGGCCGCTGCATCGATG
CCCATGAGGCGGTGCTGCGCGAGGCGGGCCTCAG
CGAGGCCCAGATCCAGACGGCGGTGCGCGTCGGC
GCCATCATCGCCTCCGCCGCGGTGGCCCTCGAGGC
CGCCGGCGCGGGTTTTCCGGAAGCTGCGGAATAA
GCGCGGCCCTGCCTCCGTCCCCGCCGCCAGGGCTT
CGGGGACGGATACCTTCAAAATTTGATCCAGATCA
AAGCCGTCCGTTGTCTCCGCGGTTTCCATTATAAC
AATTCCATTAAATCGACGGCCCGGCACACGAGGC
CGCTTCAGGGAACACCGACAGGAGGCTGATCCCT
CGCAGCCGCTTTTGCGGCAGGGAGCATGCCAATG
ACCGACGCCTTCCAGGTCCGCGTACCTCGGGTTTC
GTCGAGCACCGCCGGAGACATCGCCGCGTCATCC
ATCACCACGCGGGGCGCGCTGCCGCGCCCGGGAG
GGATGCCTGTGTCCATGTCGCGGGGGACCTCGCCC
GAGGTGGCACTCATCGGGGTCTATGAGATATCGA
AGATCCTGACGGCGCCCCGGCGCCTCGAAGTCAC
GCTCGCCAATGTGGTGAACGTGCTCTCCTCCATGC
TGCAGATGCGGCATGGCATGATCTGCATCCTCGAC
AGCGAGGGCGATCCCGACATGGTGGCCACCACCG
GCTGGACGCCTGAGATGGCGGGCCAGATCCGCGC
GCATGTGCCCCAGAAGGCCATCGACCAGATCGTC
GCCACGCAGATGCCGCTGGTGGTGCAGGACGTGA
CGGCCGATCCGCTCTTCGCCGGTCACGAGGATCTG
TTCGGCCCGCCTGAGGAGGCCACCGTCTCCTTCAT
CGGCGTGCCGATCAAGGCCGACCACCATGTGATG
GGCACCCTCTCCATCGACCGCATCTGGGACGGCAC
CGCCCGTTTCCGCTTCGACGAGGACGTGCGCTTCC
TCACCATGGTGGCCAATCTCGTCGGCCAGACCGTG
CGCCTGCACAAGCTGGTGGCGAGCGACCGCGACC
GGCTGATCGCCCAGACGCACCGCCTCGAAAAGGC
GCTGCGGGAAGAAAAATCCGGGGCCGAGCCGGAG
GTGGCCGAGGCCGCCAACGGATCCGCCATGGGCA
TCGTGGGCGATAGCCCGCTGGTGAAACGCCTGATC
GCGACCGCGCAAGTGGTCGCCCGCTCAAACTCCAC
CGTGCTGCTGCGCGGGGAGAGCGGCACCGGCAAG
GAGTTGTTCGCCCGTGCCATCCACGAACTGTCGCC
CCGCAAGGGCAAGCCCTTCGTGAAGGTGAACTGC
GCCGCCCTCCCGGAATCGGTGCTGGAATCGGAACT
GTTCGGCCATGAGAAGGGCGCCTTCACCGGTGCGC
TGAACATGCGCCAGGGCCGCTTCGAGCTGGCGCA
CGGCGGCACGCTCTTCCTTGACGAGATCGGCGAGA
TCACCCCCGCTTTCCAGGCCAAGCTGCTGCGCGTG
CTGCAGGAAGGCGAGTTCGAGCGGGTCGGCGGCA
ATCGCACGCTGAAGGTGGATGTGCGGCTCGTGTGC
GCCACCAACAAGAATCTGGAAGAGGCGGTCTCCA
AGGGCGAGTTCCGGGCCGATCTCTACTACCGCATC
CATGTGGTGCCGCTGATCCTGCCGCCGCTGCGCGA
ACGGCCGGGCGACATTCCCAAGCTCGCGAAGAAC
TTCCTCGACCGCTTCAACAAGGAAAACAAGCTCCA
CATGATGCTCTCGGCGCCGGCCATCGACGTGCTGC
GGCGCTGCTATTTCCCGGGCAACGTGCGCGAGCTG
GAGAACTGTATCCGGCGGACGGCAACGCTCGCCC
ACGATGCCGTCATCACCCCCCATGACTTCGCCTGC
GACAGCGGCCAGTGCCTCTCGGCCATGCTCTGGAA
GGGCTCGGCCCCGAAGCCTGTGATGCCGCACGTGC
CGCCGGCGCCCACGCCGCTGACTCCGCTCTCCCCT
GCTCCGCTCGCGACCGCAGCGCCCGCTGCGGCGA
GCCCGGCGCCGGCGGCCGACAGCCTGCCGGTCAC
TTGCCCCGGCACCGAGGCCTGTCCCGCGGTGCCCC
CCCGCCAGAGCGAAAAGGAGCAGTTGCTCCAGGC
CATGGAGCGCTCCGGCTGGGTGCAGGCGAAGGCC
GCGCGCCTCCTCAACCTCACGCCGCGCCAGGTGGG
TTATGCGCTGCGCAAATATGACATCGACATCAAGC
GCTTCTGAACCACGGGGCGGAGCTTGCTCCGCCCT
GTCCCGTCCGGCCGCCGATCGCGGCGCCCGCGGG
AAAGCGCCGGGCCGTTTTATGGGAGCCCCCCTCCC
ATGAGCGGTGAGCACACGGCACCGGAGCCCGCCG
CGCCGCCGGATCCGCCGGGAAGCTTCGAGGCGCT
GGACTGGCAGGGACGCCCCGTGCGCTGCGAGGAT
TGCCCGCACGAGGACATTCAGGCCATCGGCCGCTG
CGATCTCGGCAAGGTCTGCGTGGCCGACCGGCGTA
CCCGTCGGATCGACCGCTTCTTCGCCCGCGAATCC
GGAACTGGCGG

3.2. 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 gene sequences were aligned using CLUSTALW [26]. The nucleotide sequences of *nif*A gene in different nitrogen-fixing bacteria as opened in MEGA-X using MX: Alignment Explorer (FASTA). Fig. 4 illustrates the sub-window used for analysis of the alignment of conserved genes in different nitrogen-fixing organisms using MEGA-X software. The conserved regions and regions of dissimilarity were identified from multiple sequence alignment using iterative algorithms of CLUSTALW and MEGA software. Different nucleotide sequences showed consensus among different nitrogen-fixing bacteria (Fig. 4, 5). Based on the nucleotide sequence database similarity, the relatedness of different *nif*A gene sequences were compared by making the phylogenetic trees. Generated trees were viewed using TREE VIEW and best fit tree was selected out of all trees.

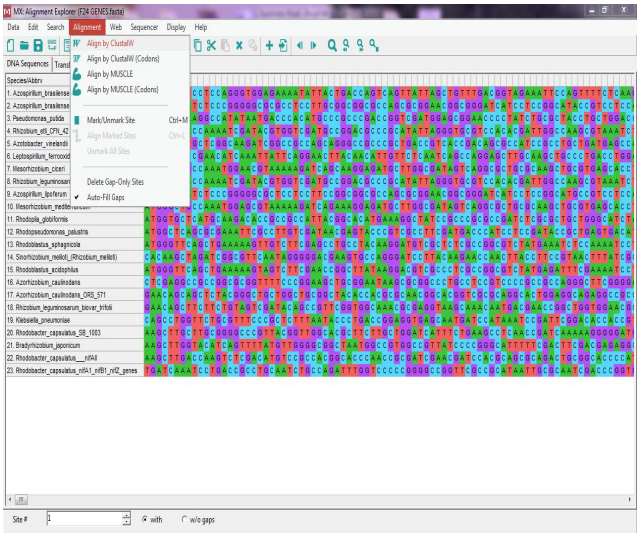

Figure 4 Interface of nucleotide sequences of *nif*A gene in different nitrogen-fixing bacteria

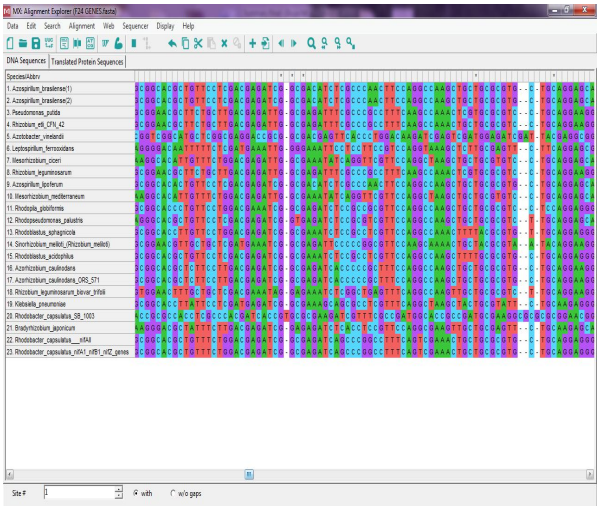

Figure 5 Alignment of nucleotide sequences of *nif*A gene in different nitrogen-fixing bacteria

Using Maximum Likelihood method, *Azorhizobium caulinodans* showed similarity with the *Rhodoblastus sphagnicola, Rhodoblastus acidophilus* and *Rhodopila globiformis* (Fig. 6). Other nodule-forming bacteria *Mesorhizobium mediterraneum, Mesorhizobium ciceri* (chick pea nodulating bacteria) and *Bradyrhizobium japonicum* (soybean nodulating bacteria) were found closely related. Other nodule-forming bacteria such as *Rhizobium leguminosarum, Rhizobium etli* (bean nodulting bacteria) and *Sinorhizobium meliloti* (alfalfa nodulting bacteria) were also distantly related. The nitrogen-fixing bacteria i.e., *Azospirillum lipoferum* and *Azospirillum brasilense*, which exist as associative symbionts were also found closely related, whereas free-living nitrogen-fixing bacteria *Klebsiella pneumoniae* and *Azotobacter vinelandii* were distantly related.

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 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. By using bootstrapping values, phylogenetic trees were also generated for *nif*A gene sequences from *Azorhizobium caulinodans* by Maximum Likelihood method. In this method, nitrogen-fixing bacteria i.e., *Azospirillum lipoferum* and *Azospirillum brasilense*, which fix nitrogen in plant tissues as associative symbionts were also found closely related to *nif*A sequences of *Azorhizobium caulinodans* (Fig. 7). *Rhodoblastus sphagnicola, Rhodoblastus acidophilus* and *Rhodopseudomonas palustris* and *Bradyrhizobium japonicum* were found closely related. *Sinorhizobium meliloti* and *Rhizobium leguminosarum* were found distantly related with reference to *nif*A sequences using Maximum Likelihood method. Free-living nitrogen-fixing bacteria i.e., *Klebsiella pneumoniae* and *Azotobacter vinelandii* as well as *Rhodobacter capsulatus* were found distantly related.

4. DISCUSSION

Nitrogen-fixing organisms, containing nitrogenase enzyme complex, belong to the kingdoms eubacteria and archaebacteria. This process of biological nitrogen fixation accounts for the majority of nitrogen transferred from the atmospheric reservoir into the biosphere. These bacteria supply fixed nitrogen to the global nitrogen cycle in diverse ecosystems and therefore, diazotrophs are present in virtually all ecosystems. The distribution pattern of microbial communities suggests that nitrogen-fixing ability is evolutionary ancient and mainly transmitted vertically with the widespread loss of function [27]. 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 [28]. Considering the importance and application of nitrogenase enzyme in agricultural field, the present study was undertaken for phylogenetic analysis of *nif*A gene sequences in *Azorhizobium caulinodans*, which fixes nitrogen in nodules formed on the root as well stem of *Sesbania* plants.

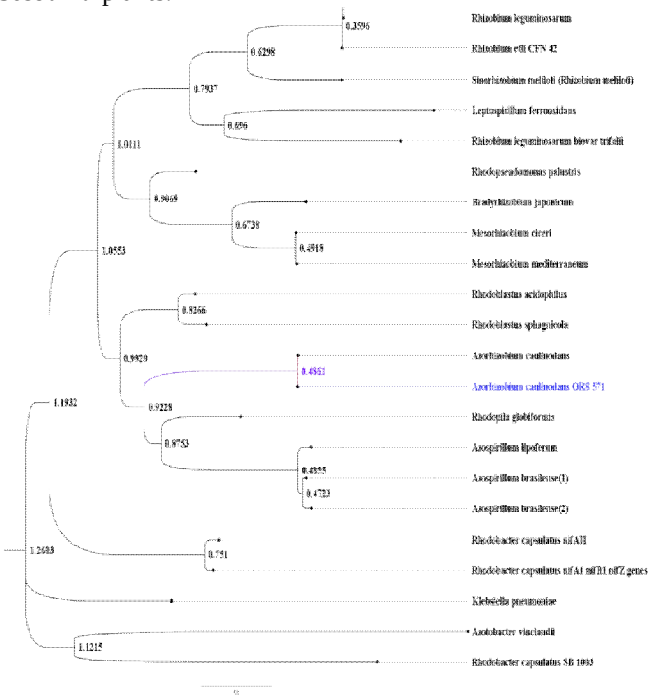

Figure 6 Phylogenetic tree of *nif*A gene by Maximum Likelihood method (without bootstrap)

Most of the characterized rhizobial strains had been derived from the limited range of cultivated legume species. Phylogenetically rhizobia are very diverse and currently include 12 genera and more than 113 species of α- and β-proteobacteria [29]. These rhizobial isolates belong to three distinct branches within the alpha-2 subgroup of *Proteobacteria* [30]. In each case, rhizobia are phylogenetically intertwined with non-symbiotic bacteria [31]. The largest branch includes the genus *Rhizobium*, which nodulates peas and clovers, and *Sinorhizobium*, which nodulates alfalfa (lucerne) and closely related to

Agrobacterium and to *Brucella*. A second branch includes the genus *Bradyrhizobium*, with species that nodulate soybean, lupin and many tropical legumes, which is closely related to *Rhodopseudomonas.* The third group includes *Azorhizobium*, which is closely related to the chemoautotroph *Xanthobacter*.

The host range of *A. caulinodans* for effective nodulation is very narrow: nitrogen-fixing nodules are formed both on the root as well as stem on *S. rostrata* and *S. punctata* [32]. *Sesbania rostrata* and other *Sesbania* species also can enter into symbiosis with other rhizobia [33], [34], including the newly described species *Sinorhizobium saheli* and *Sinorhizobium teranga* [35]. *S. saheli* and *S. teranga* belong to the group containing *Rhizobium meliloti* and *Rhizobium fredii*, which have recently been placed in the genus *Sinorhizobium*, which is phylogenetically distant from *Azorhizobium* [35].

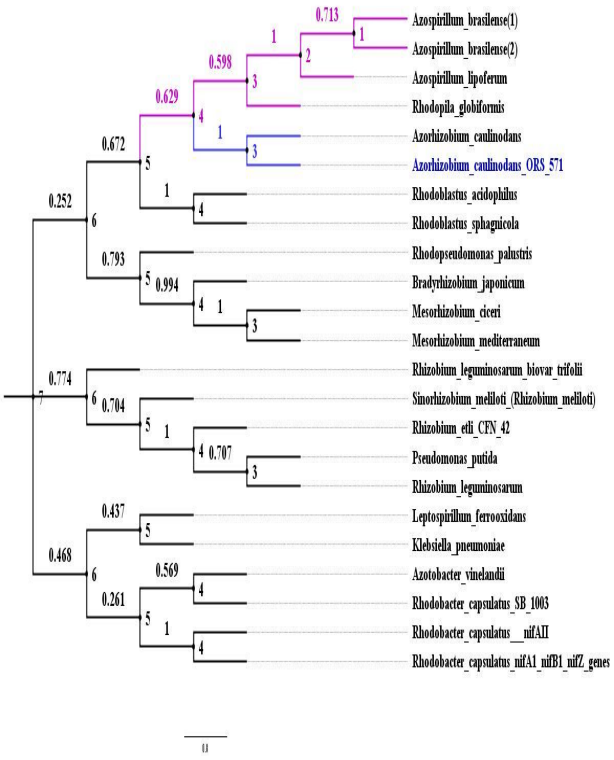

Figure 7 Neighbor-joining phylogenetic tree based on nucleotide sequences of nifA gene using Maximum Likelihood method. Bootstrap analysis was performed with 1000 cycles

The creation of a centralized, well-described and aligned *nif*A gene database of 16S ribosomal RNA (rRNA) gene sequences is the urgent need for their utility in ecology of nitrogen-fixing microorganisms [36]. In earlier taxonomic studies, Jarvis et al. [37] showed that *A. caulinodans* strain ORS571 belongs to the *Rhodopseudomonas palustris* rRNA branch of purple bacteria, but that it is quite distinct from both *Rhodopseudomonas* and *Bradyrhizobium* spp. On the other hand, *A. caulinodans* was considered as a separate genus with *Xanthobacter* as closest relative, based on numerical analysis of phenotypes, protein patterns, and DNA-DNA and DNA-rRNA hybridizations studies [33]. These *Xanthobacter* species are diazotrophic bacteria found in diverse soil habitats and in association with rice (*Oryza sativa*) roots [38], [39]. Comparison of 16S rRNA sequences indicated that *X. flavus* and *A. caulinodans* are strongly related.

The phototrophic bacterium *Heliobacterium chlorum* formed a cluster with *Desulfitobacterium hafniense*, the closest neighbour of heliobacteria based on the 16S rRNA phylogeny and two species of the genus *Geobacter* belonging to the δ-proteobacteria [40]. Thus, phylogenetic position of *Hbt. chlorum* nitrogenase may reflect an evolutionary stage of a divergence of the two nitrogenase groups, with group I consisting of the aerobic diazotrophs and group II consisting of strictly anaerobic prokaryotes. In another study, the sequence genome analysis of *Bradyrhizobium* sp. strain DOA9 showed that this strain contains the structural genes of dinitrogenase (*nif*DK) and the *nif*A regulatory gene on both the plasmid and chromosome [41]. Using *gus*A (β-glucuronidase) reporter strains, it was observed that both *nif*A genes were expressed during both the free-living and symbiotic growth stages. Furthermore, transcriptional analysis showed that NifAc and NifAp activated the expression of both chromosome and plasmid structural *nif*DK genes during symbiosis, while only NifAc activated the expression of *nif*DKc during free-living conditions.

5. CONCLUSION

Various nutrients such as nitrogen, phosphorus, potassium and zinc are required for proper growth and development of leguminous and cereal plants. These nutrients are provided to the crops mostly through application of chemical nitrogenous and phosphatic fertilizers to the soil. However, excessive and indiscriminate use of these chemical fertilizers has polluted the environment and causes various public health hazards along with slow deterioration in soil health [42] and decline in crop yield [7]. Thus, biological nitrogen fixation is of considerable economic and ecological implications, which proved to be an important component of sustainable organic farming. The number of N2-fixing plant-associated bacteria identified is still growing. Recently efforts have been made to improve nitrogen fixation with characterization of efficient nitrogen-fixing bacteria, identification and manipulation of nif genes, the genetic engineering of diazotophic bacteria along with computation modeling of the Nif proteins. The N2 fixation efficiency of the diazotrophic bacteria can be increased by manipulation of structural or regulatory nif genes of the nitrogenase enzyme complex. Thus, learning more about biological nitrogen fixation and how to improve the availability of plant-utilizable fixed nitrogen will help humanity to restore soil fertility and the health of the earth planet.

REFERENCES

[1]. Marti-Renom, M.A., Stuart, A.C., Fiser, A., Sanchez, R., Melo, F., and Sali, A., "Comparative protein structure modeling of genes and genomes", Annual Review of Biophysics and Biomolecule Structure, Vol. 29, pp. 291-298, 2000.

[2]. Sindhu, S., and Sindhu, D., "Data mining and gene expression analysis in bioinformatics", International Journal of Computer Science and Mobile Computing, Vol. 6, Issue 5, pp.72–83, 2017.

[3]. Raymond, J., Siefert, J. L., Staples, C. R., and Blankenship, R. E., "The natural history of nitrogen fixation", Molecular Biology Evolution, Vol. 21, pp. 541–554, 2004.

[4]. Kumar, R., Koul, S., Kumar, P., and Kalia, V.C., "Searching biomarkers in sequenced genomes of Staphylococcus for their rapid identification", Indian Journal of Microbiology, Vol. 56: pp. 64–71, 2016.

[5]. Singh, N., and Kashyap, S., "In silico identification and characterization of 1-aminocyclopropane-1-carboxylate deaminase from Phytophthora sojae", Journal of Molecular Modeling, Vol. 18, pp. 4101–4111, 2012.

[6]. Pramanik, K., Soren, T., Mitra, S., and Maiti, T.K., "In silico structural and functional analysis of Mesorhizobium ACC deaminase", Computational Biological Chemistry, Vol. 68, pp, 12–21, 2017.

[7]. Bohlool, B.B., Ladha, J.K., Garrity, D.P., and George, T., "Biological nitrogen fixation for sustainable agriculture: perspective", Plant and Soil, Vol. 141, Issue 1-2, pp. 1-11, 1992.

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

[9]. O'Carroll, I.P., and Dos Santos, P.C., "Genomic analysis of nitrogen fixation", Methods in Molecular Biology, Vol. 766, pp. 49-65, 2011.

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

[11]. Burris, R.H., and Roberts, G.P., "Biological nitrogen fixation", Annual Review of Nutrition, Vol. 13, pp. 317-335, 1993.

[12]. Freiberg, C., Fellay, R., Bairoch, A., Broughton, W.J., Rosenthal, A., and Perret, X., "Molecular basis of symbiosis between Rhizobium and legumes", Nature, Vol. 387, pp. 394–401, 1997.

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

[14]. Fischer, H.M., "Genetic regulation of nitrogen fixation in rhizobia", Microbiological Reviews, Vol. 58, pp. 352-386, 1994.

[15]. Ndoye, I., and Dreyfus, B., "N2 fixation by Sesbania rostrata and Sesbania sesban estimated using 15N and total N difference method", Soil Biology and Biochemistry, Vol. 20, pp. 209–213, 1988.

[16]. Ladha, J. K. Pareek, R.P., and Becker, M., "Stem-nodulating legume- Rhizobium symbiosis and its agronomic use in low land rice", Advances in Soil Science, Vol. 20, pp. 147–192, 1992.

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

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

[19]. Dixon, R., "Stranglehold on transcriptional activator by its partner regulatory protein - case of NifL-NifA two component regulatory system", In: SGM Symposium 61: Signals, switches, regulons and cascades: control of bacterial gene expression (Hodgson D.A., and Thomas C.M., eds.). Cambridge University Press, pp. 213-230, 2002.

[20]. Lakhani, J., Khuteta, A., Choudhary, A., and Harwani, D., "Hierarchical clustering-based algorithms and in silico techniques for phylogenetic analysis of rhizobia", In: Hansen A., Choudhary D., Agrawal P., and Varma A. (eds.) Rhizobium Biology and Biotechnology. Soil Biology, Springer, Cham. Vol. 50, pp. 185-214, 2017.

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

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

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

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

[25]. Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Sayers, E.W., "GenBank", Nucleic Acids Research, Vol. 38, pp. D46–D51, 2010.

[26]. Higgins, D., Thompson, J.D., Higgins, D.G., and Gibson, T.J., "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.

[27]. Hartmann, L.S., and Barnum, S.R., "Inferring evolutionary history of Mo-dependent nitrogen fixation from phylogenetic studies of nifK and nifDK", Journal of Molecular Evololution, Vol. 71, pp. 70-85, 2010.

[28]. Dos Santos, P.C., Fang, Z., Mason, S., Setuba, J.C., and Dixon, R., "Distribution of nitrogen fixation and nitrogenase-like sequences amongst microbial genomes", BMC Genomics, Vol. 13, pp. 162, 2012. doi: 10.1186/1471-2164-13-162

[29]. Sawada, H., Kuykendall, L.D., and Young, J.M., "Changing concepts in systematics of bacterial nitrogen-fixing legume symbionts", Journal of General and Applied Microbiology, Vol. 49, pp. 155–179, 2003.

[30]. Willems, A., "Taxonomy of rhizobia: overview", Plant and Soil, Vol. 287, pp. 3–14, 2006.
[31]. Moulin, L., Munive, A., Dreyfus, B., and Boivin-Masson, C., "Nodulation of legumes by members of β-subclass of Proteobacteria", Letters Nature, Vol. 411, pp. 948–950, 2001.
[32]. Boivin, C., Ndoye, I., Molouba, F., de Lajudie, P., Dupuy, N., and Dreyfus B., "Stem nodulation in legumes: diversity, mechanisms and unusual characters", Critical Reviews in Plant Sciences, Vol. 16, pp. 1–30, 1997.
[33]. Dreyfus, B., Garcia, J.L., and Gillis, M., "Characterization of Azorhizobium caulinodans gen. nov, sp. nov, stem-nodulating nitrogen-fixing bacterium isolated from Sesbania rostrata", International Journal of Systemic Bacteriology, Vol. 38, pp. 89-98, 1988.
[34]. Rinaudo, G., Orenga, S., Fernandez, M.P., Meugnier, H., and Bardin, R., "DNA homologies among mem.bers of the genus Azorhizobium and other stem- and root-nodulating bacteria isolated from the tropical legume Sesbania rostrata", International Journal of Systematic Bacteriology, Vol. 41, pp. 114-120, 1991.
[35]. de Lajudie, P., Willems, A., Pot, B., Dewettinck, D., Maestrojuan, G., Neyra, M., Collins, M.D., Dreyfus, B.L., Kersters, K., and Gillis, M., "Polyphasic taxonomy of rhizobia. Emendation of the genus Sinorhizobium and description of Sinorhizobium meliloti comb. nov., Sinorhizobium saheli sp. nov. and Sinorhizobium teranga sp. nov.", International Journal of Systematic Bacteriology, Vol. 44, pp. 715–733, 1994.
[36]. Cole, J. R., Wang, Q., and Cardenas, E. et al., "Ribosomal database project: improved alignments and new tools for rRNA analysis", Nucleic Acid Research, Vol. 37, pp. D141–D145, 2009.
[37]. Jarvis, B.D.W., Gillis, M., and DeLey, J., "Intra- and intergeneric similarities between the ribosomal ribonucleic acid cistrons of Rhizobium and Bradyrhizobium species and some related bacteria", International Journal of Systemic Bacteriology, Vol. 36, pp. 129-138, 1986.
[38]. Reding, H.K., Hartel, P.G., and Wiegel, J., "Effect of Xanthobacter, isolated and characterized from rice roots, on growth on wetland rice", Plant and Soil Vol. 138, pp. 221–229, 1991.
[39]. Wiegel, J., "The genus Xanthobacter", In: The Prokaryotes A Handbook on the Biology of Bacteria: Ecophysiology, Isolation, Identification, Applications. 2. III. Balows, A., Trüper, H.G., Dworkin, M., Harder, W., and Schleifer, K-H. (eds.), Berlin: Springer-Verlag. pp. 2365–2383, 1992.
[40]. Enkh-Amgalan, J., Kawasaki, H., and Seki, T., "Molecular evolution of nif gene cluster carrying nifI1 and nifI2 genes in Gram-positive phototrophic bacterium Heliobacterium chlorum", International Journal of Systematic and Evolutionary Microbiology, Vol. 56, pp. 65-74, 2006.
[41]. Wongdee, J., Boonkerd, N., Teaumroong, N., Tittabutr, P., and Giraud, E., "Regulation of nitrogen fixation in Bradyrhizobium sp. strain DOA9 involves two distinct NifA regulatory proteins that are functionally redundant during symbiosis but not during free-living growth", Frontiers in Microbiology, Vol. 9, pp. 1644, 2018.
[42]. Bockman, O.C., "Fertilizers and biological nitrogen fixation as sources of plant nutrients: Perspectives for future agriculture", Plant and Soil, Vol. 194, pp. 11-14, 1997.

AUTHORS

Divya Sindhu completed her B. Tech. degree in Computer Science and Engineering from Kurukshetra University, Kuruksheta in 2012. She received her M. Tech. degree in Computer Science and Engineering from Guru Jambheshwar University of Science and Technology, Hisar in 2015. Her areas of research 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 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.

Saurabh Sindhu successfully completed his B.C.A. and M.C.A. degrees from Guru Jambheshwar University of Science and Technology, Hisar. His areas of research interest include data mining, data security, E-agriculture, 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 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.

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.