ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 8, Issue 3 , March 2021

Computational Biology: Use of NifA Protein Amino Acid Sequences of *Azorhizobium* strain for Phylogenetic Analysis among Nitrogen-fixing Organisms

Divya Sindhu*, Neeru Singh Redhu, Saurabh Sindhu , S. K. Yadav

Department of Computer Science and Engineering, Shri Jagdishprasad Jhabarmal Tibrewala University, Jhunjhunu - 333001, Rajasthan, India
Department of Molecular Biology, Biotechnology and Bioinformatics, CCS Haryana Agricultural University, Hisar – 125004, Haryana, India
Department of Computer Science and Engineering, Shri Jagdishprasad Jhabarmal Tibrewala University, Jhunjhunu - 333001, Rajasthan, India
Department of Computer Science and Engineering, Shri Jagdishprasad Jhabarmal Tibrewala University, Jhunjhunu - 333001, Rajasthan, India

ABSTRACT: Soil nutrients are the most vital components for efficient crop production. To improve soil fertility and crop yield in sustainable agriculture, recently nitrogen-fixing bacteria are used as biofertilizers. These bacteria use nitrogenase enzyme to fix atmospheric nitrogen into ammonia for utilization by plants. Various genes and proteins involved in nitrogen fixation have been characterized. NifA protein of nodule-forming *Azorhizobium caulinodans* plays an important role in activation of nitrogen fixation (*nif*) genes under nitrogen-limited and microaerobic conditions. Databases of various genes/proteins involved in nitrogen fixation are available in GenBank and computational tools are used for phylogenetic and molecular evolution studies in data mining. In the present study, amino acid sequences of NifA protein from 15 different nitrogen-fixing bacterial species were retrieved from NCBI GenBank and phylogenetic analysis was performed using Maximum Likelihood method. NifA protein of *Azorhizobium caulinodans* showed maximum aminoacid similarity with *Bradyrhizobium diazoefficiens* followed by *Rhizobium etli* and *Sinorhizobium fredii*, where as other nodule-forming bacteria i.e. *Rhizobium leguminosarum* bv. *trifolii* and *Sinorhizobium meliloti* were found quite distantly spaced. Associative symbionts including *Azospirillum brasilense, Azospirillum lipoferum, Herbaspirillum seropodicae* and *Pentoea agglomerans* were found quite distantly related to *Azorhizobium caulinodans.* Similarly, free-living nitrogen-fixing bacteria *Klebsiella pneumoniae* and *Azotobacter vinelandii* were also found distantly related.

KEYWORDS: NifA protein, Amino acids, *Azorhizobium caulinodans*, NCBI GenBank, Phylogenetic analysis, Nitrogen-fixing bacteria, Biofertilizers

I. INTRODUCTION

Intensive farming practices adopted by farmers involve the use of costly inputs such as chemical fertilizers and pesticides to provide sufficient food for increasing world population. However, excessive and injudicious application of fertilizers has caused environmental pollution and public health hazards. Therefore, use of nitrogen-fixing organisms as biofertilzers has emerged recently as cost effective and eco-friendly technology to reduce the application of chemical fertilizers [1]. Currently, biofertilizers make up the main component of integrated nutrient management thereby leading to sustainability [2]. These nitrogen-fixing organisms possess nitrogenase enzyme, which causes reduction of atmospheric inert nitrogen to plant utilizable ammoniacal form. This process of biological nitrogen (N_2) fixation is sporadically distributed among both eubacteria and methanogenic archaea [3, 4]. Among different N_2-fixing systems, legume-*Rhizobium* symbiotic association has been reported to reduce about 70-80% of the total of 17.2×10^7 tonnes biologically fixed nitrogen per year [5]. The genes involved in nitrogen fixation (*nif* genes) have been identified in the free-living *Klebsiella pneumoniae* and symbiotic rhizobia using genomic analysis and transcriptional profiling [6, 7].

Due to rapid developments in genomics and proteomics involving novel sequencing technologies, large amounts of biological data have been generated. The detailed nucleotide sequence information of the various *nif* genes is now available in biological databases i.e., the NCBI GenBank [8]. Sophisticated computational and bioinformatical analyses

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 8, Issue 3 , March 2021

are required using data mining (DM) approaches to draw conclusions from these biological data [9, 10]. Computational predictions obtained by applying phylogenetic profile method usually measure functional interactions between proteins across multiple genomes [11, 12]. Since, biological functions result from the interactions of proteins, therefore understanding the network of biological linkages utilizing functional genomics and proteomics information is becoming a major thrust area currently [13].

The varied applications of phylogenetic tools, algorithms and use of *in silico* methods to study phylogeny of rhizobia may be considered as a highly reliable and important technique in biological sciences [14] and may replace wet experiments for prediction of evolutionary relationships between two microbial species in a laboratory. Till now, the current understanding of nitrogenase diversity has been based largely on phylogenetic analyses of nitrogenase structural genes i.e., *nif*H and *nif*D [15, 16]. Besides these structural genes in N_2-fixing (diazotrophic) organisms, NifA protein plays a major role in transcriptional activation and expression of nitrogenase structural genes in response to high oxygen and ammonia availability [17]. In order to understand important microorganism-host interactions, a computational approach was employed in this study to explore the phylogenic relatedness and evolutionary relationships of NifA protein using *Azorhizobium caulinodans* among other nitrogen-fixing organisms.

II. LITERATURE SURVEY

The ability to fix nitrogen is widely distributed among archaea and bacteria. Use of a minimum set of six genes encoding for structural and biosynthetic components of nitrogenase enzyme, namely *Nif*HDK and *Nif*ENB were proposed as a new criterion for computational prediction of nitrogen fixation [18]. Using this criterion, 149 diazotrophic species were identified from fully sequenced genomes, including 82 known diazotrophs and 67 species not known to fix nitrogen. The phylogenetic position of phototrophic bacterium *Heliobacterium chlorum* was the same in the NifH, NifD, NifK, NifE and NifN trees [19]. *Hbt. 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. 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. Using novel computational methods, about 263 novel proteins were identified potentially associated with the symbiosis interactome in the symbiosis of *Sinorhizobium meliloti* with its alfalfa host plants [20].

Various studies demonstrated that nitrogenase enzyme first 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 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 and with a maximum of 20 genes in obligate aerobes *Azotobacter vinelandii* [21]. Lau et al. [22] performed community phylogenetics and phylogeography of microorganisms living in extreme environments. The taxonomic patterns indicated the effect of historical, geographical and environmental factors on microbial functions. Frank et al. [23] suggested that a novel approach is needed for utilizing NifH amino acid sequences to rapidly classify into well-organized phylogenetic clusters. The utility of this novel sequence binning approach revealed a marine–terrestrial distinction in community composition. 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 to fix nitrogen during both free-living and symbiotic growth [24].

III. METHODOLOGY

Efficient computational and bioinformatics tools afford novel opportunities for understanding biological information from genomic and proteomic databases [9]. Phylogeny and phylogenetic trees give a picture of the evolutionary history among species, individuals or genes [25]. In the present study, phylogeny of NifA protein of the nodule-forming bacteria was done by taking *Azorhizobium caulinodans*, which make nodules on *Sesbania*.

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 8, Issue 3 , March 2021

A. Retrieval of NifA protein sequences in different nitrogen-fixing bacteria

In order to have complete information regarding the features of macromolecular structures, the protein data base (PDB) allows a wide spectrum of queries through data integration. Database similarity search tool, FASTA, was applied that work on heuristic method of database searching. 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 was used as the search point [27]. In the present study, amino acid sequences of NifA protein from 15 different nitrogen-fixing and nodule-forming bacterial strains were retrieved from NCBI GenBank and Uniprot KB Database. Phylogenetic analysis with other nitrogen-fixing bacteria and nodule-forming rhizobial species, was done by taking NifA protein sequences of *Azorhizobium caulinodans*, which fixes nitrogen in root and stem nodules of *Sesbania*.

B. Phylogenetic analysis of NifA protein among different nitrogen-fixing bacteria using computational software

Partial sequences for amino acid sequences of NifA protein of different nitrogen-fixing and nodule-forming rhizobia were removed from retrieved sequence datasets, obtained from NCBI GenBank and UniprotKB databases. 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. Molecular Evolutionary Genetics Analysis (MEGA) computer software (i.e., MEGA-X) was used for statistical analysis of molecular evolution and for constructing phylogenetic trees. Based on the amino acid sequence database similarity, the relatedness of different amino acid sequences of NifA protein were compared by making the phylogenetic trees. Consensus trees were constructed for all sequences with or without bootstrapped method using both softwares. The number of replications (iterations) used to construct the phylogenetic tree were taken as 1000.

C. Phylogenetic trees construction

Phylogenetic trees were built by character based methods using Maximum Likelihood (ML) method [28]. Maximum Likelihood 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 method is well suited to the analysis of distantly related sequences [29]. The best fit tree was obtained by with 1000 bootstrapping replications in MEGA (Fig. 1). Values above nodes represented bootstrap values. 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. Phylogenetic relationships of genes or organisms usually are presented in a tree-like form with a root.

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 8, Issue 3 , March 2021

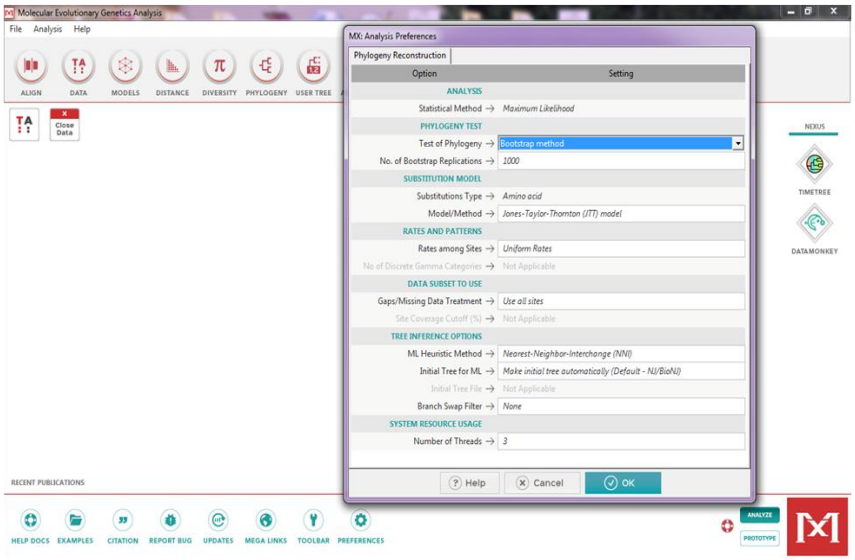

Fig. 1. Construction of Phylogenetic trees of amino acids with bootstrap method. The number of replications (iterations) used to construct the Phylogenetic Tree is 1000. The Statistical method used is Maximum Likelihood.

IV. EXPERIMENTAL RESULTS

NifA protein plays major role in transcriptional activation of nitrogen fixation (*nif*) structural genes [17]. Sequences of NifA protein varies between 519 aminoacids in *Rhizobium leguminosarum* to 605 aminoacids in *Bradyrhizobium japonicum*, whereas *Mesorhizobium ciceri* possesses only 352 aminoacids. NifA protein consists of three domains: amino-terminal domain, a central domain and carboxy-terminal domain. Central domain consists of 240 aminoacids, which are sufficient by itself for activation [30]. Its C-terminal domain with a helix-turn-helix motif binds with the upstream activator sequences and contributes towards transcriptional activation.

A. Sequence retrieval of nitrogen fixation NifA protein in nitrogen-fixing and nodule-forming rhizobia

In this study, amino acid sequences of NifA protein from 15 different nitrogen-fixing bacterial strains were retrieved in FASTA format for computational analysis from NCBI GenBank [8, 27] or UniprotKB Database [31], respectively. 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 performed for protein sequences. Particular query used for filtering the NifA protein in nitrogen-fixing bacteria in NCBI (Fig. 2).

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 8, Issue 3 , March 2021

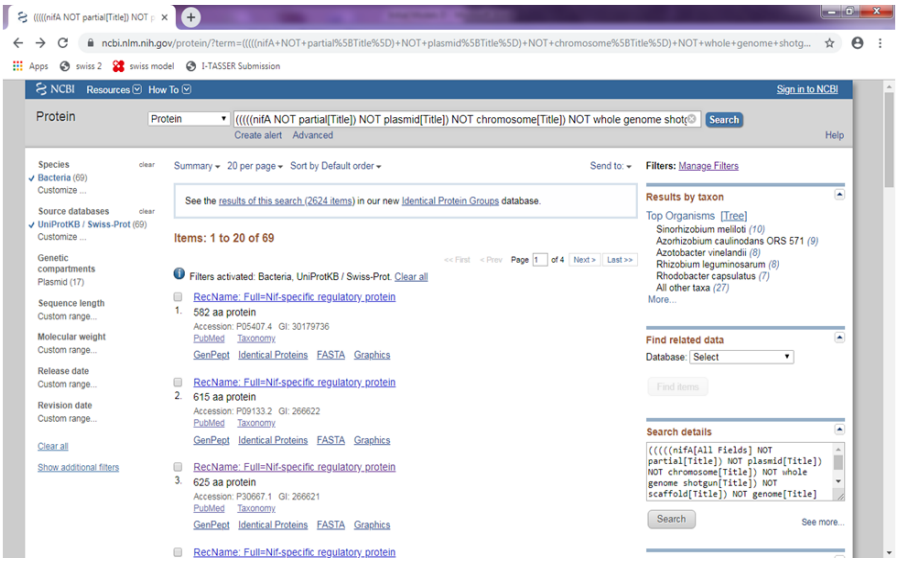

Fig. 2. Sequence retrieval of NifA protein from UniProtKB database

For biosynthesis of various proteins in organisms, twenty distinct amino acids are utilized by living cells. Amino acids join together 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 used in synthesis of proteins 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 aminoacid sequences of NifA protein in different nitrogen-fixing organisms were opened in FASTA format in NOTEPAD.

Amino acid sequences of following eight nitrogen-fixing bacteria are provided i.e. *Azospirillum brasilense, Azotobacter vinelandii, Azorhizobium caulinodans* ORS 571, *Bradyrhizobium diazoefficiens, Herbaspirillum seropodicae, Klebsiella pneumoniae, Rhizobium leguminosarum* bv. *trifolii* and *Rhodobacter capsulatus.* These bacteria were isolated from differentat habitats and range from microaerobic to aerobic life style, free-living, associative diazotrophs to nodule-forming symbiotic bacteria.

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

MPMTDAFQVRVPRVSSSTAGDIAASSITTRGALPRPGGMPVSMSRGTSPEVALIGVYEISKILTAPRRLEVTLANVVNVL
SSMLQMRHGMICILDSEGDPDMVATTGWTPEMAGQIRAHVPQKAIDQIVATQMPLVVQDVTADPLFAGHEDLFGPPEEAT
VSFIGVPIKADHHVMGTLSIDRIWDGTARFRFDEDVRFLTMVANLVGQTVRLHKLVASDRDRLIAQTHRLEKALREEKSG
AEPEVAEAANGSAMGIVGDSPLVKRLIATAQVVARSNSTVLLRGESGTGKELFARAIHELSPRKGKPFVKVNCAALPESV
LESELFGHEKGAFTGALNMRQGRFELAHGGTLFLDEIGEITPAFQAKLLRVLQEGEFERVGGNRTLKVDVRLVCATNKNL
EEAVSKGEFRADLYYRIHVVPLILPPLRERPGDIPKLAKNFLDRFNKENKLHMMLSAPAIDVLRRCYFPGNVRELENCIR
RTATLAHDAVITPHDFACDSGQCLSAMLWKGSAPKPVMPHVPPAPTPLTPLSPAPLATAAPAAASPAPAADSLPVTCPGT
EACPAVPPRQSEKEQLLQAMERSGWVQAKAARLLNLTPRQVGYALRKYDIDIKRF

>***Bradyrhizobium diazoefficiens*** **USDA 110**

MLHIPSSSERPASQPEPERAPPGEPSHESALAGIYEISKILNAPGRLEVTLANVLGLLQSFVQMRHGLVSLFNDDGVPEL
TVGAGWSEGTDERYRTCVPQKAIHEIVATGRSLMVENVAAETAFSAADREVLGASDSIPVAFIGVPIRVDSTVVGTLTID
RIPEGSSSLLEYDARLLAMVANVIGQTIKLHRLFAGDREQSLVDKDRLEKQTVDRGPPARERKQLQAHGIIGDSPALSAL
LEKIVVVARSNSTVLLRGESGTGKELVAKAIHESSVRAKRPFVKLNCAALPETVLESELFGHEKGAFTGAVSARKGRFEL
ADKGTLFLDEIGEISPPFQAKLLRVLQEQEFERVGSNHTIKVDVRVIAATNRNLEEAVARSEFRADLYYRISVVPLLLPP

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 8, Issue 3 , March 2021

LRERRSDIPLLAREFLRKFNSENGRSLTLEASAIDVLMSCKFPGNVRELENCIERTATLSAGTSIVRSDFACSQGQCLST
TLWKSTSYGKTDPAAPMQPVPAKSIIPLAETAPPPQAVCEPGSLAPSGTVLVSGARMADRERVVAAMEKSGWVQAKAARL
LGLTPRQVGYALRKYGIEIKRF

>***Azospirillum brasilense***

MPGAMRQSTSNLELLTIYEVSKILGSSLDLQQTLREVLRALAYQLQMHRGRVYLVGEDNVLRLVAANGLSNEAAAQIEFR
DGEGITGRILKTGMPAVVPNLAEEPLFLNRTGGREDLDEQVASLVGVPIKAAGVVVGVLTIDRISDEGPQGHFGSDVRFL
TMVANLIGQTVRLHRTVAEERRFMMRETFRMQKELRPVAAPINDVVCTSPNMLEVMAQVHRVAPFKSTVLIRGESGTGKE
LIARAIHNMSPRKDAPFIRVNCAALPESLLESELFGHEKGAFTGAQKDHKGRFELASGGTLFLDEIGDISPNFQAKLLRV
LQEQEFERVGGSKTIKTDVRLICATNLNLEEAVGHGKFRADLYFRINVVTIHLPPLRERRQDIGPLARHFVAKFAKDNGM
ALVMEDEALEVLNRCTWPGNVRELENCIERAATQSRDGIIRTESLSCSLNLCNSSVLFQYRTLGASVGGLAPSMGPGAIN
RVPPGRPGGPAAANAPKTPAMPAPVPEPAGAAAARGRPARRVVPRPLAGLRRRPAGGSGPPDPACPCPSRAPLPPQAPPP
SPAAAPPPAAEVPLDEPESGSLRDRLLWAMERTGWVQAKAARLLGMTTRQVSYALRKYNIEIKRF

>***Rhizobium leguminosarum* bv. *trifolii***

MPKSTVRPGAEIVGESAALKEVLEIAQIVARSNSPVLLRGESGTGKEFFAKLIHDSSSRHEKPFVKLNCAALSAGVLESE
LFGHEKGAFTGATSQKEGRFELAHGGTLLLDEIGEISAEFQAKLLRVLQEGELERVGGTRTLKVNVRLVCATNKDLETAV
AAGEFRADLYYRINVVPITLPPLRQRDGDIPRLAQKFLQRFNRENGRSLSFAPATLDILSKCEFPGNIRELQNCTQRTAT
LARSDVIVPQDLACEQGRCYSPILKKAVAEQVGKGAIHGLARGETESMGQPCDVGVFAAETVMGQSGLIGRERLEQAMAT
AGWVQAKAARLLGRTPRQVGYSLRRHGIERKVF

>***Klebsiella pneumoniae***

MIHKSDSDTTVRRFDLSQQFTAMQRISVVLSRATEASKTLQEVLSVLHNDAFMQHGMICLYDSQQEILSIEALQQTEDQT
LPGSTQIRYRPGEGLVGTVLAQGQSLVLPRVADDQRFLDRLSLYDYDLPFIAVPLMGPHSRPIGVLAAHAMARQEERLPA
CTRFLETVANLIAQTIRLMILPTSAAQAPQQSPRIERPRACTPSRGFGLENMVGKSPAMRQIMDIIRQVSRWDTTVLVRG
ESGTGKELIANAIHHNSPRAAAAFVKFNCAALPDNLLESELFGHEKGAFTGAVRQRKGRFELADGGTLFLDEIGESSASF
QAKLLRILQEGEMERVGGDETLRVNVRIIAATNRHLEEEVRLGHFREDLYYRLNVMPIALPPLRERQEDIAELAHFLVRK
IAHSQGRTLRISDGAIRLLMEYSWPGNVRELENCLERSAVLSESGLIDRDVILFNHRDNPPKALASSGPAEDGWLDNSLD
ERQRLIAALEKAGWVQAKAARLLGMTPRQVAYRIQIMDITMPRL

>***Azotobacter vinelandii***

MNATIPQRSAKQNPVELYDLQLQALASIARTLSREQQIDELLEQVLAVLHNDLGLLHGLVTISDPEHGALQIGAIHTDSE
AVAQACEGVRYRSGEGVIGNVLKHGNSVVLGRISADPRFLDRLALYDLEMPFIAVPIKNPEGNTIGVLAAQPDCRADEHM
PARTRFLEIVANLLAQTVRLVVNIEDGREAADERDELRREVRGKYGFENMVVGHTPTMRRVFDQIRRVAKWNSTVLVLGE
SGTGKELIASAIHYKSPRAHRPFVRLNCAALPETLLESELFGHEKGAFTGAVKQRKGRFEQADGGTLFLDEIGEISPMFQ
AKLLRVLQEGEFERVGGNQTVRVNVRIVAATNRDLESEVEKGKFREDLYYRLNVMAIRIPPLRERTADIPELAEFLLGKI
GRQQGRPLTVTDSAIRLLMSHRWPGNVRELENCLERSAIMSEDGTITRDVVSLTGVDNESPPLAAPLPEVNLADETLDDR
ERVIAALEQAGWVQAKAARLLGMTPRQIAYRIQTLNIHMRKI

>***Rhodobacter capsulatus* SB1003**

MTDQQSRPASPRRRSTQSIADRLALDALYEIAKTFAAAPDPVAEVPQIFNVLSSFLDLRHGVLALLAEPGEGAGVNPYVI
AATAFQRSPEAPAADVLPDAVARIVFRSGVPFVSFDLAAEFGAEAVPKRLRDAGQTLIAVPLRDPERSHFVLGVLAAYRS
HDHNRSGFSDADVRVLTMVASLLEQALRFRRRIARDRERALEDTRRMLQTVTEQRGPAAPVSLDGIVGSSPAIAEVVAQI
KRVASTRMPVLLRGESGTGKELFARAVHAQSPRAKGPFIRVNCAALSETLLESELFGHEKGAFTGATALKKGRFELADGG
TLFLDEIGEISPAFQSKLLRVLQEGEFERVGGAKTIKVDTRIVAATNRDLEDAVARGQFRADLYFRICVVPIVLPPLRNR
KSDIKPLAQLFLDRFNKQNATNVKFAADAFDQICRCQFPGNVRELENCVNRAAALSDGAIVLAEELACRQGACLSAELFR
LQDGTSPIGGLAVGRVITPTVRVSAPPPEPAPAPEPAPEAPPREEVPLRTKTAQLSREELLRALESAGWVQAKAARLLGM
TPRQIAYALQKFEIELRKI

>***Herbaspirillum seropedicae***

MATILDDRSVNLELVTIYEISKILGSSLDLSKTLREVLNVLSAHLETKRVLLSLMQDSGELQLVSAIGLSYEEFQSGRYR
VGEGITGKIFQTETPIVVRDLAQEPLFLARTSPRQSQDGEVISFVGVPIKAAREMLGVLCVFRDGQSPSRSVDHEVRLLT
MVANLIGQTVRLYRSVAAERQQLQEEKRQLSRQLQGKYKLDNVIGISKAMQEVFAQVHQSAPSRSTMLLRGESGTGKEVI
ARAIHYLSPRKDGPFIKVNCAALSETLLESELFGHEKGAFTGAQGERKGRFELAHGGTLFLDEIGEISPAFQAKLLRVLQ
EREFERVGGSRSIKVDVRLVTATNRDLEKAVAKGEFRADLYYRINVVSIFIPPLRERREDIPYLVEHFLEKFRVENQRAM

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 8, Issue 3 , March 2021

VAMSPQAMKVMMNCYWPGNVRELENCVERTATMMRGDLITEVHFSCQQNKCLTKVLHEPGQQQPVVVVPLERISAPYGAI
FAEWDGQGQATGAAPPTSERERLIWAMEQCGWVQAKAARALNISPRQMGYALQKFNIEVKKF

B. Alignment of amino acid sequences

The amino acid and gene sequences were aligned using CLUSTALW [32]. The amino acid sequences of NifA protein in different nitrogen-fixing bacteria as opened in MEGA-X using MX: Alignment Explorer (Protein FASTA) is shown in Fig. 3. Different aminoacids shows consensus in different nitrogen-fixing bacteria. Conserved amino acid residues of NifA protein in different nitrogen-fixing organisms using MEGA-X software are demonstrated. Different aminoacids shows consensus in different nitrogen-fixing bacteria. Conserved regions were found in the alignment from 274-526 and 651-696 amino acids.

C. To generate the phylogenetic tress among different nitrogen-fixing and nodule-forming bacteria based on the sequence similarity index

In this study, the level of similarity was searched among the amino acid sequences of different NifA proteins. Conserved region and region of dissimilarity among amino acid sequences were identified using multiple sequence alignment. Based on the nucleotide sequence database similarity, the relatedness of different NifA amino acid sequences were compared by making the phylogenetic trees using Maximum Likelihood method (Fig. 3). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Evolutionary analyses were conducted in MEGA-X. The final trees were edited and viewed in Figtree.

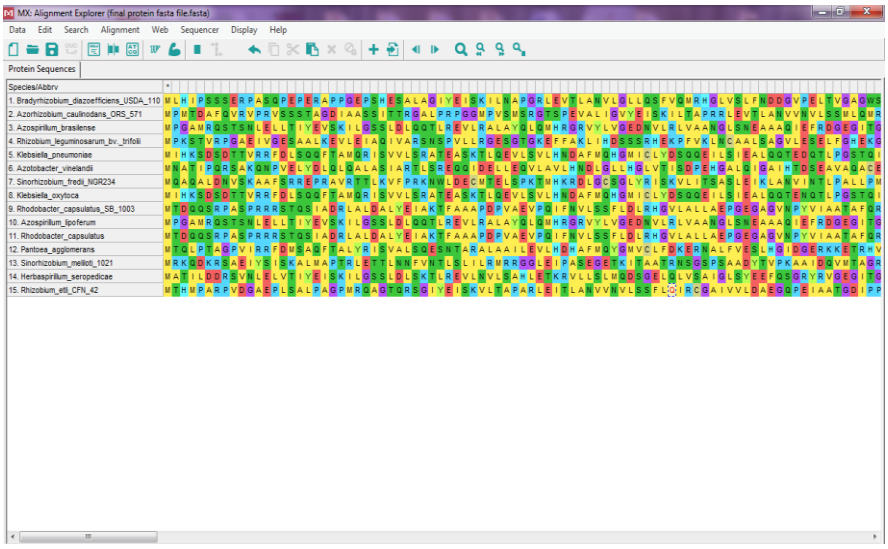

Fig. 3. Amino acid sequences of nifA protein in different nitrogen-fixing bacteria opened in MEGA-X software

By using Maximum Likelihood approach, *Azorhizobium caulinodans* showed maximum aminoacid similarity with *Bradyrhizobium diazoefficiens* followed by *Rhizobium etli* and *Sinorhizobium fredii* (Fig. 4). Diazotrophic bacteria which fix nitrogen in the plant tissues as associative symbionts such as *Azospirillum brasilense, Azospirillum lipoferum, Herbaspirillum seropodicae* and *Pentoea agglomerans* were found quite distantly related to *Azorhizobium caulinodan.* Using the Maximum Likelihood method, free-living nitrogen-fixing bacteria *Klebsiella pneumoniae, Klebsiella oxytoca, Azotobacter vinelandii* as well as associative bacteria *Herbaspirillum seropodicae, Azospirillum brasilense* and *Azospirillum lipoferum,* were found distantly related (Fig. 5). However, *Sinorhizobium fredii, Rhizobium leguminosarum* bv. *trifolii* and *Sinorhizobium meliloti* were found quite distantly spaced.

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 8, Issue 3 , March 2021

V. DISCUSSION

Taxonomic data obtained from 16S ribosomal RNA (rRNA) genes using various culture-independent approaches has been mainly used for evaluations of microbial biogeography of surface habitats. Phylogenetic studies indicate the evolutionary history of genes and species, and therefore, is currently one of the most important subjects in molecular evolution. Reliable phylogenies of diverse genes and proteins from different organisms help us to understand the mechanisms of evolution as well as the history of organisms. Thus, phylogenetic tree is a two dimensional representation of relatedness among various biological species.

Bacteria containing nitrogenase enzyme complex supply fixed nitrogen to the global nitrogen cycle. Due to this critical role in nitrogen cycle, diazotrophs are present in virtually all ecosystems, varying from aerobic soils (e.g., *Azotobacter* species), the ocean surface layer (*Trichodesmium*) and specialized nodules in legume roots (*Rhizobium*). Phylogenetic classification of *nif*H gene sequences has already been demonstrated as an essential step in diazotroph community analysis [23, 33]. The inoculation of *Azorhizobium caulinodans* has been found to form nitrogen-fixing nodules on both the root as well as stem of the legume crop *Sesbania rostrata* and increased levels of plant biomass were exhibited [34, 35]. Due to dual nodulation topology on the host *Sesbania*, *A. caulinodans* is phylogenetically separated from other rhizobia and offered a unique system for investigating its interaction with its host.

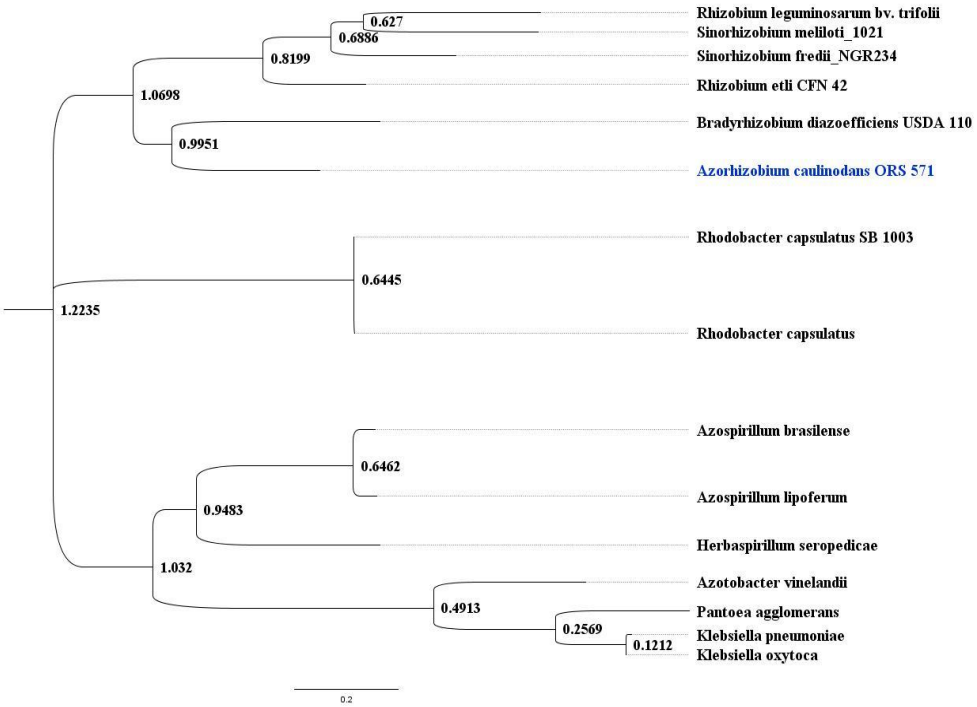

Fig. 4. Phylogenetic tree of NifA protein using Maximum Likelihood method

Most of the characterized rhizobial strains belong to three distinct branches within the alpha-2 subgroup of *Proteobacteria* [36]. In each case, rhizobia are phylogenetically intertwined with non-symbiotic bacteria [37]. 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 *nif*A gene database of 16S ribosomal RNA (rRNA) gene sequences has been utilized in ecology of nitrogen-fixing microorganisms [38, 39]. In earlier taxonomic studies, Jarvis et al. [40] showed that *A. caulinodans* strain ORS571

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 8, Issue 3 , March 2021

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 as well as DNA-rRNA hybridization studies [41]. de Lajudie et al. [42] reported that *Sinorhizobium saheli* and *Sinorhizobium 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*. Similarly, comparison of 16S rRNA sequences indicated that *Xanthobacter flavus* and *A. caulinodans* are strongly related [43].

VII. CONCLUSION AND SCOPE

Various nutrients such as nitrogen, phosphorus and potassium are required for proper growth and development of leguminous and cereal plants. These nutrients are provided by farmers mostly through application of chemical fertilizers to the soil. However, indiscriminate use of these fertilizers has polluted the environment and causes various health hazards [44] and a decline in crop yield [45]. Due to harmful side effects of these agrochemicals, scientific community is involved in identification of beneficial microorganisms, which could be used in the soil as biofertilizers. Szeto et al. [46] hypothesized that increasing the NifA production, which is the transcriptional activator of other *nif* genes, could enhance the expression of the whole N_2-fixing system. Subsequently, some *R. meliloti* strains having enhanced *nif*A gene products were found to cause 7-15% increase in alfalfa plant biomass in comparison to wild-type parent in green house conditions. In this study, phylogenetic analysis of NifA protein using *Azorhizobium caulinodans* showed its maximum aminoacid similarity with nodule-forming strains of *Bradyrhizobium diazoefficiens* followed by *Rhizobium etli* and *Sinorhizobium fredii*. Free-living and associative nitrogen-fixing bacteria were found quite distanly related to *A. caulinodans.* Additional research is required to understand the dynamic interactions between nitrogen-fixing organisms, crop plants and the environment to increase crop yield and quality [47]. Furthermore, streamlined protocols and knowledge-based designs have to be developed for use of nitrogen-fixing bacteria for improving crop productivity to feed the ever-increasing human population.

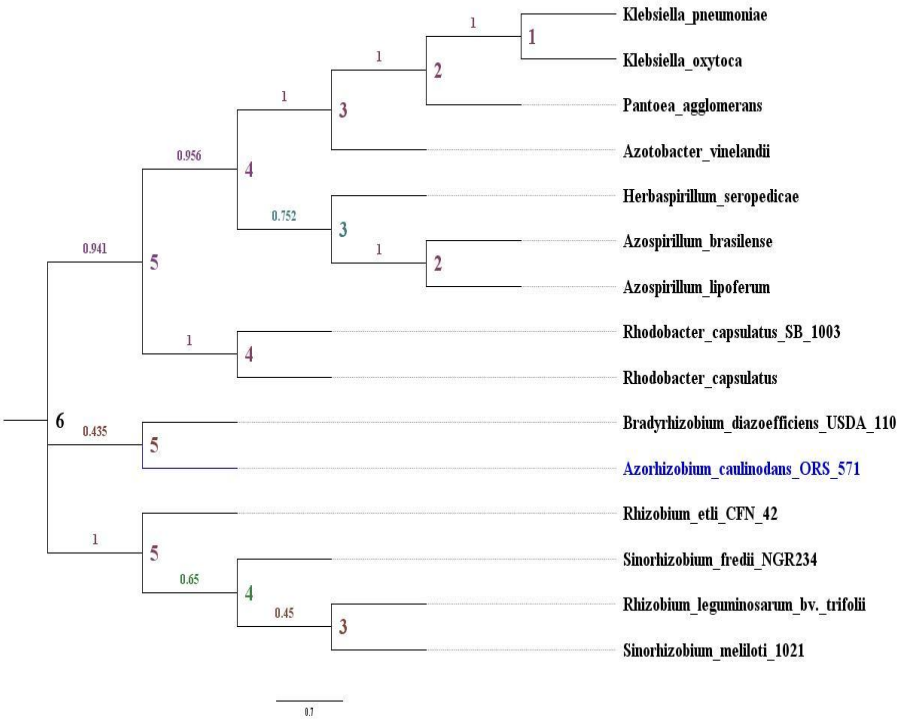

Fig. 5. Phylogenetic tree of NifA protein using bootstrap method (no. of iterations = 1000) by Maximum Likelihood method

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 8, Issue 3 , March 2021

REFERENCES

[1]. Kour, D., Rana, K.L., Yadav, A.N., Yadav, N., Kumar, M. et al., "Microbial biofertilizers: Bioresources and eco-friendly technologies for agricultural and environmental sustainability". Biocatalysis and Agricultural Biotechnology, vol. 23, 2020, pp. 101487.

[2]. Prasad, S., Malav, L.C., Choudhary, J., Kannojiya, S., Kundu, M., Kumar, S., Yadav, A.N., "Soil microbiomes for healthy nutrient recycling". In: Current trends in microbial biotechnology for sustainable agriculture, Yadav, A.N. et al. (eds.). Springer Nature, Singapore Pte Ltd. 2021. doi.org/10.1007/978-981-15-6949-4_1

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

[4]. 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 and Hall. 1992, pp. 43–86.

[5]. Ishizuka, J., "Trends in biological nitrogen fixation research and application". Plant and Soil, vol. 141, 1992, pp. 197-209.

[6]. 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, 1992, pp. 763-831.

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

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

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

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

[11]. Marcotte, E.M., Pellegrini, M., Ng, H.L., Rice, D.W., Yeates, T.O., Eisenberg, D., "Detecting protein function and protein-protein interactions from genome sequences". Science, vol. 285, 1999, pp. 751-753.

[12]. Pellegrini, M., Marcotte, E.M., Thompson, M.J., Eisenberg, D., Yeates, T.O., "Assigning protein functions by comparative genome analysis: protein phylogenetic profiles". Proceedings National Academy Sciences USA, vol. 96, 1999, pp. 4285-4288.

[13]. Enright AJ, Iliopoulos I, Kyrpides NC, Ouzounis CA, Butland G, Peregrín-Alvarez JM, Li J, Yang W. et al., "Interaction network containing conserved and essential protein complexes in *Escherichia coli*". Nature, vol. 433, 2005, pp. 531-537.

[14]. Sindhu, D., "Data mining and *in silico* modeling of nitrogen fixation NifA protein in nodule-forming rhizobia using Bioinformatics tools", Ph. D. thesis submitted to Shri Jagdishprasad Jhabarmal Tibrewala University, Jhunjhunu, Rajasthan, India, 2020, pp. 251.

[15]. 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, 2003, pp. 539–554.

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

[17]. 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, 2002, pp. 213-230.

[18]. 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, 2012, pp. 162. doi: 10.1186/1471-2164-13-162

[19]. Enkh-Amgalan, J., Kawasaki, H. and Seki, T., "Molecular evolution of *nif* gene cluster carrying *nif*I1 and *nif*I2 genes in Gram-positive phototrophic bacterium *Heliobacterium chlorum*". International Journal of Systematic and Evolutionary Microbiology, vol. 56, 2006, pp. 65-74.

[20]. Rodriguez-Llorente, I., Caviedes, M.A., Dary, M., Palomares, A.J., Cánovas, F.M. and Peregrín-Alvarez, J.M., "Symbiosis Interactome: computational approach reveals novel components, functional interactions and modules in *Sinorhizobium meliloti*". BMC Systems Biology, vol. 3, 2009, pp. 63. doi: 10.1186/1752-0509-3-63

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

[22]. Lau, M.C.Y., Cameron, C., Magnabosco, C., Brown, C.T., Schilkey, F., Grim, S., Hendrickson, S., Pullin, M. 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, 2014, pp. 1-17.

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

[24]. 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, 2018, pp. 1644. doi: 10.3389/fmicb.2018.01644

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

[26]. 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, 1997, pp. 3389.

[27]. 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, 2010, pp. D5–D16.

[28]. Felsenstein, J., "Evolutionary trees from DNA sequences: maximum likelihood approach". Journal of Molecular Biology, vol. 17, 1981, pp. 368.

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

[30]. Huala, E. and Ausubel, F.M., "Central domain of *Rhizobium meliloti* NifA is sufficient to activate transcription from *R. meliloti nif*H promoter". Journal of Bacteriology, vol. 171, issue 6, 1989, pp. 3354-3365.

[31]. Berman, H., Henrick, K., Nakamura, H. and Markley, J.L., "Worldwide Protein Data Bank (wwPDB): ensuring single, uniform archive of PDB data". Nucleic Acids Research, vol. 35, 2007, pp. D301–D303.

[32]. 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, 1994, pp. 4673-4680.

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 8, Issue 3 , March 2021

[33]. Sindhu, S., Sindhu, D., Yadav S.K., "Data mining and phylogenetic analysis of NifH protein of *Azospirillum* strain among nitrogen-fixing bacteria using bioinformatics tools". International Journal of Computer Science and Engineering, vol. 9, issue 1, 2021, pp. 1-10,.

[34]. Ndoye, I. and Dreyfus, B., "N_2 fixation by *Sesbania rostrata* and *Sesbania sesban* estimated using ^{15}N and total N difference method". Soil Biology and Biochemistry, vol. 20, 1988, pp. 209–213.

[35]. 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, 1992, pp. 147–192.

[36]. Willems, A., "Taxonomy of rhizobia: overview". Plant and Soil, vol. 287, 2006, pp. 3–14.

[37]. Moulin, L., Munive, A., Dreyfus, B. and Boivin-Masson, C., "Nodulation of legumes by members of β-subclass of Proteobacteria". Letters Nature, vol. 411, 2001, pp. 948–950.

[38]. Alexandre, A., Laranjo, M., Young, J.P.W. and Oliveira, S., "DnaJ is useful phylogenetic marker for alpha-proteobacteria". International Journal of Systematic Evolution in Microbiology, vol. 58, 2008, pp. 2839-2849.

[39]. 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, 2009, pp. D141–D145.

[40]. Jarvis, B.D.W., Gillis, M. and DeLey, J., "Intra- and inter-generic similarities between the ribosomal ribonucleic acid cistrons of *Rhizobium* and *Bradyrhizobium* species and some related bacteria". International Journal of Systematic Bacteriology, vol. 36, 1986, pp. 129-138.

[41]. Dreyfus, B.L., 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, 1988, pp. 89-98.

[42]. 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, 1994, pp. 715–733.

[43]. Rainey, F.A. and Wiegel, J., "16S ribosomal DNA sequence analysis confirms the close relationship between the genera *Xanthobacter*, *Azorhizobium* and *Aquabacter* reveals a lack of phylogenetic coherence among *Xanthobacter* species". International Journal of Systematic Bacteriology, vol. 46, 1996, pp. 607–610.

[44]. Newbould, P., "Use of nitrogen fertilizer in agriculture: where do we go sustainablely and ecologically"? Plant and Soil, vol. 115, 1989, pp. 297–311.

[45]. 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, 1992, pp. 1-11.

[46]. Szeto, W., Kwiatkowski, R., Cannon, F.C. and Ronson, C.W., "Enhancement of symbiotic nitrogen fixation in *Bradyrhizobium japonicum*". In: Abstracts of fifth international symposium on molecular genetics of plant-microbe interactions. Interlaken, Switzerland, 1990, pp. 152.

[47]. Zhang, J., Cook, J., Nearing, J.T., Zhang, J., Raudonis, R., Glick, B.R., Langille, M.G.I. and Cheng, Z., "Harnessing the plant microbiome to promote the growth of agricultural crops". Microbiological Research, vol. 245, 2021, pp. 1-14.