Bio
Technology

Symbiotic Nitrogen Fixation-A Bioinformatics Perspective

[1]S. Sur, [2]A.K. Bothra and [1]A. Sen
[1]Department of Botany, NBU Bioinformatics Facility, University of North Bengal,
Siliguri-734013, West Bengal, India
[2]Department of Chemistry, Raiganj University College, Raiganj-733134, West Bengal, India

Abstract: The important element nitrogen can be fixed by symbiotic as well as asymbiotic nitrogen fixers to make it available as nutrient in the ecosystem. Symbioses with higher plants offer an ecological niche for a particular microorganism to fix nitrogen. Although there have been significant advances in nitrogen fixation research, the development of bioinformatics has opened a new window for providing novel understanding into the mechanism. The accessibility of complete genome sequences of symbiotic diazotrophs became a stepping stone for gaining insights into the working principle of bacterial cell. Studies on diazotrophs genomes exposed new evidences pertaining to evolution, structure, interactions between plants and microbes and diversity amongst them. Detection of a number of symbiotic genes associated with nitrogen fixation has strengthened functional genomic research. Accessibility of new software tools for genomic and proteomic analysis have accelerated research in the areas of codon usage, proteome analysis, molecular modeling and phylogenetic analysis of nitrogen fixation genes. In the present review we highlight the discovery of nitrogen fixation, research trends in nitrogen fixation in the pre-genomic era, the implications of bioinformatics research on nitrogen fixation, post-genomic research on BNF with special reference to synonymous codon usage analysis, proteome analysis, analysis of the molecular phylogeny of nitrogen fixation using nucleotide triplet based condensed matrix methods and the ambiguities in alignment based methods and use of homology modeling techniques in the structural biology of nitrogenases.

Key words: Nitrogen fixation, symbiosis, codon usage, proteome, molecular phylogeny

INTRODUCTION

Nitrogen is stable, inert element comprising 78% of the Earth's atmosphere. It is one of the indispensable elements that limit the primary production on which life forms depend for their daily requirements. It is a component of the proteins required for plant, animal and human life (Rosswall, 1976). Inspite of its abundance a small percentage of it is capable of entering into the biogeochemical cycle. Approximately 0.001% of the nitrogen occurs in the biosphere and the rest remains in the lithosphere (Stevenson, 1972). Nitrogen occurs in molecular form in the atmosphere. Merely 0.04% of the nitrogen occurs in compounds that are accessible to living beings. In the terrestrial system, nitrogen accounts for 4% of the biomass and 96% of soil organic matter (Rosswall, 1976). An important feature of our planet ecosystem is the nitrogen cycle since nitrogen availability affects the rate of major ecosystem processes, like primary production and decomposition. It is a mechanism by which nitrogen is converted into different chemical forms. Both biological and non-biological processes are associated with these transformations. Nitrogen cycle consists of nitrogen fixation, assimilation, mineralization, nitrification and denitrification (Rosswall, 1976). The process of conversion of molecular nitrogen to ammonia is known as nitrogen fixation. Some fixation occurs during lightning strikes; however most of the part is played a group of free-living (e.g., *Azotobacter*) or symbiotic bacteria (*Rhizobium, Bradyrhizobium, Frankia* etc.). This fixed nitrogen is the primary source of available nitrogen in nature and is an essential step distributing the supply of this nutrient in the ecosystem (Saikia and Jain, 2007). Majority of plants and microorganisms incorporate ammonium salts as source of nitrogen either through glutamate dehydrogenase or glutamine synthetase/ glutamate synthase cycle (Miflin and Lea, 1977) by the process of assimilation. By the process of mineralization ammonium is released as molecular nitrogen. Some free living organisms like *Nitrobacter, Nitrosomonas* etc. utilize ammonium or nitrite compounds as an energy source (Ritchie and Nicholas, 1972). These are nitrifying organisms and the process is called nitrification (Rosswall, 1976). Denitrification reduces nitrates back to

Corresponding Author: Arnab Sen, NBU Bioinformatics Facility, Department of Botany, University of North Bengal,
Siliguri-734013, West Bengal, India

the largely inert nitrogen gas (N_2) (Rosswall, 1976). Some bacterial species like *Pseudomonas* and *Clostridium* use nitrate as an electron acceptor in place of oxygen during respiration in anaerobic conditions and are denitrifiers. These denitrifiers complete the nitrogen cycle.

However, nitrogen fixation also occurs by industrial synthesis, UV radiation and lightening. During industrial synthesis, Haber-Bosch process is used to produce nitrogenous fertilizers which are energy consuming and complicated (Ladha and Reddy, 1995). Nitrogen fertilizers are applied in agriculture in either nitrate (NO^{3-}) or ammonium (NH^{4+}) forms for enhancing crop production in order to meet the demands of the ever increasing world population. The chief constituent of the Green revolution of the 20th century was heavy dependency on chemical fertilizers, pesticides and herbicides obtained from fossil fuels. Bumb and Baanante (1996) sketched an annual increase of about 1.2% in the use of nitrogenous fertilizer 1990 to 2020. For the production of wheat, rice and maize alone 42 million tons of nitrogenous fertilizers are used annually on a global scale (Saikia and Jain, 2007). Approximately 25% of the nitrogen from nitrogenous fertilizers is lost by leakage, volatilization, denitrification and other factors resulting in high economic loss accompanied by severe environmental pollution. The high presence of nitrate and nitrite ions in water, food has resulted in diseases like cancer in new borns, respiratory problems, kidney failure etc. (Saikia and Jain, 2007). The ecosystem gets damaged owing to high levels of nitrate and ammonium ions in cultivated soils resulting in plant toxicity (Saikia and Jain, 2007). In these circumstances, global concerns over the ill effects of chemical fertilizers, in addition to their rising costs for farmers in third world countries, it is necessary to reap the benefit of Biological Nitrogen Fixation (BNF) based methodologies for cost effectiveness, maximizing output and obliterating the ill effects of environmental pollution.

Biological Nitrogen Fixation (BNF) is the mechanism of reduction of atmospheric dinitrogen to a metabolically active form exclusively by microorganisms. Biological nitrogen fixation is directly proportional to agricultural sustainability (Bohlool *et al.*, 1992). An understanding of the factors controlling BNF systems in the field is vital for the support and successful adoption in large scale. Gone are the concepts of trial-and-error experiments in agricultural sciences (Bohlool *et al.*, 1992). It is necessary to explore the possibilities for improving biological nitrogen fixation and its use by farmers on a global scale.

Early civilizations like that of the Egyptians knew the importance of legumes for maintaining soil fertility. The significance of crop rotation, green manuring and intercropping systems was practiced by major civilizations like the Romans and the Greeks. However, it was only in the 19th century that the role of microorganisms in nitrogen accumulation and thus crop improvement came to be recognized. It was reported that Baussingault (Rao, 1988), Lawes and Gilbert (1895) in the 19th century worked on the diverse capacities of legumes and cereals to accumulate nitrogen but failed to understand the mechanism. However, the onus of the discovery fell upon two German scientists, who identified that root nodules of leguminous plants were responsible for conversion of atmospheric nitrogen to ammonia. It was the Dutch scientist Martinus Beijerinck, who isolated and cultured the microorganisms from root nodules of leguminous plants in 1888 (Rao, 1988). Beijerinck primarily named it *Bacillus radicicola* (Rao, 1988) and proved that it helps in formation of nodules in roots of leguminous plants. The modifications in the culture media took place over time and the nodulating bacterium was easily isolated and identified to be *Rhizobium*. The culture and isolation of nodulating bacteria from leguminous plants by Beijerinck opened the possibility of characterizing similar organisms in different parts of the world. Besides, a number of free-living nitrogen fixing microorganisms like *Clostridium pasteurianum* (Winogradsky, 1893), *Azotobacter chroococcum, Beijerinckia* (Rao, 1988), *Klebsiella pneumoniae, Azospirillium* etc. were discovered.

NITROGEN FIXING MICROORGANISMS

Some microorganisms enter into symbiotic association with a number of leguminous as well as non-leguminous plants. Symbioses with higher plants offer an ecological niche for the particular microbe to fix nitrogen (Mylona *et al.*, 1995). The rhizobia groups of bacteria consisting of the genera like *Rhizobium, Bradyrhizobium, Sinorhizobium, Azorhizobium* and *Mesorhizobium* with the members of the legume family (Mylona *et al.*, 1995). Different species of *Rhizobium* generally forms association with peas, lucerne, clover and beans while *Bradyrhizobium* and *Azorhizobium* forms associations with soyabeans and *Sesbania* respectively (Schlegel, 1991). It has been reported that most of the legume symbionts are host specific (Mylona *et al.*, 1995). Since the discovery of nitrogen fixation legume-rhizobia association has been the most studied relationship owing to its importance in improving soil fertility (Rao, 1988). Symbiotic association is not limited to the legumes but to a number of non-legumes. The most significant amongst them are the actinorhizal plants-*Frankia* association. This association involves a number of plants like *Casuarina, Hippophae, Alnus, Myrica,* etc. belonging to different families (Benson and Silvester, 1993). Some cereals like

rice, wheat, maize, millets, grasses etc. have symbiotic association with microorganisms accomplishing the process of nitrogen fixation (Saikia and Jain, 2007). Prominent among these microbes are *Azospirillium*, *Acetobacter* and *Azoarcus* (Saikia and Jain, 2007). Besides, the *Anabaena-Azolla* association (Bohlool *et al.*, 1992), *Nostoc–Gunnera* association (Mylona *et al.*, 1995) can fix a substantial amount of nitrogen. It has been reported (Willey, 1979) that in semi-arid tropics legumes play a leading role compared to humid tropics and intercropping system has been practiced with cereals (Fujita *et al.*, 1992). In temperate climates, soybeans (Keyser and Fudi, 1992) and forages (Ledgard and Steele, 1992) are widely used. Isolation of microorganisms from surface-sterilized roots led to the discovery of nitrogen fixing endophytes (Dobereiner, 1992; Baladani and Baldani, 2005). Cycads in association with cyanobacterial species can fix nitrogen (Costa and Lindblad, 2002).

In the rhizobia and *Frankia* the root nodules are the sites for nitrogen fixation. An important feature of the root nodules in leguminous plants is the existence of a unique red pigment called leghaemoglobin. It helps in the oxygen uptake by terminal oxidases and boosts ATP production for nitrogenase activity (Schlegel, 1991). Nodule formation is not only limited to the roots but in stems of *Aeschynomeme* and *Sesbania* and leaves of *Ardisia*, *Pavetta* and *Psychotria* (Rao, 1988). However, bacteria associated with leaf nodules do not fix nitrogen. Compared to legume symbiotic microorganisms the first isolation of *Frankia* a non legume symbiont strain was achieved from nodules of *Comptonia peregrina* in 1978 (Callaham *et al.*, 1978). *Frankia* has two distinctive developmental structures vesicles and spores crucial for its survival (Lechevalier, 1994). Vesicles are the sites for nitrogen fixation, while spores contained in are the reproductive structures in *Frankia*. In symbiotic nitrogen fixation, the establishment of the bacterium inside the host root and nodule development is a complex process and involves the processes like recognition and infection of the host root, differentiation of the root nodules, proliferation of the bacteria and transformation of the bacteroids in nodules (Schlegel, 1991).

Free living microorganisms pass independent life and fix nitrogen. Aerobic free living microbes include *Azotobacter*, *Beijerinckia*, *Azonomas* etc., facultative anerobes include *Klebsiella*, anaerobic bacteria like *Clostridium acetobutylicum*, *C. pasteurianum*, photosynthetic bacteria like *Chlorobium*, *Rhodospirillium* etc. (Schlegel, 1991). It has been established (Rao, 1988) that among cyanobacteria both heterocystous and non heterocystous forms, *Anabaena*, *Nostoc, Trichodesmium, Lyngbya, Plectonema* etc. fix

nitrogen (Schlegel, 1991). These occur in wide ranging habitats and ecological niches. Methanogenic Archaea gave a wider angle to the field of biological nitrogen fixation. Zinder and Daniels were the first to discover nitrogen fixation in different methanogenic Archaea (Murray and Zinder, 1984; Belay *et al.*, 1984). In cyanobacteria a light and thick walled cell structure called heterocyst is the point where nitrogen fixation takes place. However, in members like *Lyngbya, Plectonema* etc. where heterocyst is absent nitrogen fixation occur in internally organized cells (Schlegel, 1991).

NITROGENASE

The most important enzyme associated with the nitrogen fixing mechanism is nitrogenase. It is a quite complex biological molecule that reduces dinitrogen to ammonia at appropriate temperature and pressure (Peters and Szilagyi, 2006). Rao (1988) reported that Carnahan and his co-workers in the year 1960 were the pioneers in carrying out the reduction of dinitrogen to ammonia in cell free extracts of *Clostridium pasteurianum*. This discovery ushered the concept of universality of nitrogenase in free-living as well as symbiotic nitrogen fixing systems. Mo-nitrogenase is the conventional nitrogenase, harboring a prosthetic group with molybdenum. *Azotobacter* and some photosynthetic dizaotrophs, bear extra type of nitrogenase whose cofactor houses vanadium or only iron (Newton, 2007). The conventional nitrogenase is an equilibrium mixture of molybdo-ferro protein and iron protein in the ratio of 1:2. Two FeMoCo are attached to α subunits of the MoFe protein. Besides, 4Fe- 4S clusters comprise of some other prosthetic groups. The P-clusters are covalently linked to cysteine residues of molybdo-ferro protein forming a bridge between α and β subunits (Rubio and Ludden, 2008). Alexander (1977) stated that all nitrogen fixers had alike proteins, signifying that a similar but not matching reaction sequence. Consequently the molecular weights in the two components differ. In the later part of the 20th century studies on the enzymology and structures of nitrogenase became appropriate since, nitrogenase catalysis reactions involved biochemical processes like protein-protein interactions, signal transduction and electron-transfer reactions. The studies concerning the enzymology of nitrogenases were hampered due to the dearth of the establishment of a complete mechanism for reduction of dinitrogen with respect to the structure and properties of the nitrogenase proteins (Howard and Rees, 1996). The establishment of the three-dimensional structures for the nitrogenase proteins in 1992 and their metal clusters considerably changed the research on

nitrogen fixation (Georgiadis *et al.*, 1992; Kim *et al.*, 1993; Einsle *et al.*, 2002). The establishment of the structure paved the way for developments in the areas of biochemistry, spectroscopy, biophysics, model chemistry in describing structural features of the nitrogenase proteins, their functional mechanism and capacity to catalyse nitrogen fixation under ambient conditions by various workers (Burgess and Lowe, 1996; Rees and Howard, 2000; Lawson and Smith, 2002; Seefeldt *et al.*, 2004).

GENETICS OF NITROGEN FIXATION

Various techniques like mutations, deletion mapping, cloning vectors etc. have facilitated the identification of genes associated with nitrogen fixation. Legume, non legume and free living nitrogen fixers have a set of genes which are responsible for effective nodulation and nitrogen fixation. These are the *nod, nol, noe, nif, fix* and some hydrogenase genes. The work on the genetics of nitrogen fixation was first started in *Klebsiella oxytoca* M5a1 and first ever detailed organization of *nif* genes were reported in this organism (Arnold *et al.*, 1988). A number of studies (Zheng *et al.*, 1998; Merrick and Edwards, 1995; Dixon and Kahn, 2004; Hu *et al.*, 2007; Rubio and Ludden, 2008) have established that core *nif* genes like *nifH, nifD nifK, nifY, nifB, nifQ, nifE, nifN, nifX, nifU, nifS, nifV, nifW, nifZ* are essential for nitrogen fixation. On the basis of mutational studies the natures of different *nif* gene products were determined. Studies confirmed that *nif* HDK encodes nitrogenase while *nif* LA had regulatory functions. In cyanobacteria especially *Anabaena* sp. 7120 it was established that the organization of *nif* HDK differed from that of *Klebsiella pneumoniae.* Rao (1988) reported that it was Johnston and his co-workers who discovered the presence of nodulation genes in a plasmid of *Rhizobium leguminosarum* and mutation of those genes rendered them useless. Later on studies (Schultz and Kondorosi, 1998; Perret *et al.*, 2000) ascertained that *nod, nol* and *noe* genes produce nodulation signals. The interplay of different *nod* genes, triggering of the creation of root nodule, signaling cascades and development of nodule meristem were reported by a number of researchers (Yang *et al.*, 1999; Long, 2001; Geurts and Bisseling, 2002). Sprent and James (2007) reviewed the evolution of different nodule structures and infection types in leguminous plants. Studies (White *et al.*, 2007) concerning nodule physiology and nodule functioning specified that the bacteria in bacteroids forms are surrounded in plant membranes resulting in the formation of symbiosomes. Ammonia produced in bacteroids is transported to the plant all the way through the symbiosomes by amino acid flux system (White *et al.*, 2007). Besides, alanine is transported. The origin of nodulation genes in leguminous plants are clouded in mystery however research is on in this field. Like the rhizobia, *Azosprillium* includes a megaplasmid and sequences similar to *nod* genes (Elmerich, 1984). *Frankia* houses a number of *nif* genes but researchers failed to spot *nod* genes in *Frankia* (Cérémonie *et al.*, 1998). However, groundwork on *Frankia* genomes exposed some putative *nod*-like genes which did not emerge in organized clusters and failed to detect the *nod* A gene (Franche *et al.*, 2009). Rao (1988) reported that *fix* genes were identified on the chromosomes as well as plasmids of *Rhizobium* and were transferable. These genes are irregular in free-living bacteria (Gubler and Hennecke, 1986) and assist in electron transfer to the nitrogenase.

RESEARCH ON NITROGEN FIXATION

The discovery of nitrogen fixing organisms throughout the globe, studies concerning their gene products as well as the development of molecular biology helped BNF research enter a new stage where basic research combined with latest techniques. BNF research in the late 20th century mainly focused on exploration and application of diazotrophs resources, mechanisms of ammonium suppression, oxygen sensitivity of *nif* gene expression, regulation between plant and microorganisms during symbiosis, nature and synthesis of nodulation factor, biochemistry and functions of nitrogenase and increase in nitrogen fixation efficiency for crop productivity (Franche *et al.*, 2009). Advances in gene cloning, PCR amplifications and sequencing of DNA ushered a new way of identifying diazotrophs. Zehr and McReynolds (1989) were the pioneers in developing a number of primers for amplification of *nif* H fragments from the environment. It enabled the estimation of the diversity of diazotrophs in the rhizosphere without going through the complex process of bacterial isolation, cultural studies biochemical studies etc. (Hamelin *et al.*, 2002; Roesch *et al.*, 2008). Stoffels *et al.* (2001) reported that studies on 16S rRNA genes design of specific probes in the later part of the 20th century.

BEGINNING OF INTERDISCIPLINARY RESEARCH AND DAWN OF BIOINFORMATICS

It was in the year 1986 that the Department of Energy (DOE), USA and National Institutes of Health (NIH) started the Human Genome Project (HGP) and became one of the most happening experiments in the late 20th

century. The objective of the project was to identify all of the genes in humans and craft a database containing the information (Ideker *et al.*, 2001). A number of other genome projects also started in major industrialized countries of Europe and Japan. The scientists were at first skeptical about the HGP, since huge amount of money had to be spent that would hamper basic research and as the project moved scientists worried about the massive amount of data and its interpretation (Bloom, 2001) In the year 2000, Prof. Collins of the National Human Genome Research Institute and Prof. Craig Venter of Celera Genomics appeared in a press conference and stated that they had achieved what was thought impossible and published the draft sequence of human genome (Wade, 2000). That was the beginning and it opened the flood-gates for other genome sequencing projects. Gradually sequences of mouse, rat, worms and yeast were completed (Miller *et al.*, 2004). Kyrpides (1999) reported that in the end of the 20th century, there were 24 complete genomes that included 16 bacterial, 6 archaeal and 2 eukaryotic genomes and currently there are more than two thousand genomes available in public databases. The large numbers of genomes resulted in the generation of huge amount of information concerning the genetic nature of biological organisms spanning different kingdoms, groups and lineages etc. Bloom (2001) proposed that the greatest problem appeared to be the interpretation of underlying information from genomes leading to materialization of the new science of bioinformatics. Bioinformatics revolutionized the science of biology and directed it towards a more holistic approach compared to the reductionism visible in molecular biology research in the late 20th century (Bloom, 2001). Now an organism is viewed as a system comprising of the information associated with genes and proteins that are responsible for maintaining day to day functions and networks of regulations that spell out how gene expression occur (Ideker *et al.*, 2001). In 1995, the publication of the complete genome of *Haemophilus influenzae* marked the beginning of another revolution in the field of bioinformatics (Fleischmann *et al.*, 1995). Currently there are about two thousand bacterial and archaeal genomes in the public domain. The publication of huge amount of sequence data helped in the development of high end computers, smart computing tools, for large-scale annotation, functional classification of the proteins (Searls, 2000) and development of specific databases (Birney *et al.*, 2002) for availability to the broad scientific community. As the science of bioinformatics developed, computation became cheaper and cheaper and was duly complemented with the growth of Internet since the late 90s (Perez-Iratxeta *et al.*, 2007). High throughput tools

greatly developed in the beginning of the 21st century as genetic data became a gold mine for researchers (Perez-Iratxeta *et al.*, 2007). New software started developing for more efficient and comprehensive analysis of the genomes, proteomes and proteins. The stage was set and bioinformatics had become the leading science of the 21st century.

POST GENOMIC ERA AND BNF

The amalgamation of the knowledge of plant physiology, biochemistry, genetics and molecular biology gave idea about the understanding of the mechanism of nitrogen fixation in pre-genomic era. Nevertheless, the knowledge remained erratic and discrete (Benedito *et al.*, 2006). Things changed with the accessibility of complete genome sequences of symbiotic as well as non-symbiotic diazotrophs. Knowledge of the whole genome became the stepping stone in understanding the working principle of the bacterial cell (Puhler *et al.*, 2004). *Mesorhizobium loti* was the first sequence of a symbiotic bacterium and it was followed by *Sinorhizobium meliloti* (Puhler *et al.*, 2004). The completion of the genomes of *Rhizobium leguminosarum bv. viciae* (Young *et al.*, 2006), *Rhizobium etli* (Gonzalez *et al.*, 2006), *Bradyrhizobium* strains (Kaneko *et al.*, 2002), *Frankia* strains (Normand *et al.*, 2007) and sequences for a number of free-living diazotrophs spanning different habitats and ecological niches (Franche *et al.*, 2009) bolstered nitrogen fixation research. The studies on the genomes exposed new evidences pertaining to evolution and structure, interactions between plants and microbes and diversity amongst the diazotrophs. The breakthrough in unearthing the gene maneuvering propagation of nodule primordia opened a new dimension in functional genomics of rhizobia-legume symbiosis (Gresshoff, 2003). The detection of a number of symbiotic genes associated with nitrogen fixation has strengthened functional genomic research. Tools like DNA macro- and microarrays have been applied for studying expression at transcription level in *Sinorhizobium* and are being applied in other rhizobia (Puhler *et al.*, 2004). Genome sequences of *Frankia* yielded significant information regarding their evolutionary histories and linked the disparity in size with the biogeographic history of the host plants of each strains (Normand *et al.*, 2007). Transcriptomic and proteomic investigations of the sequenced genomes provided important information regarding rhizobial-legume symbioses (Puhler *et al.*, 2004). The analysis of the genomes of *Azoarcus* and *Glucanoacetobacter* is expected to start comparative genomic studies on endophytic as well as endosymbiotic associations

(Puhler *et al.*, 2004). MacLean *et al.* (2007) reported that genome studies have proved that the architecture of the genome content is guided by the lifestyle of legume symbionts. Since, legumes are fundamental part of agriculture and a significant component of sustainable agriculture, sequencing projects of legumes viz. *Lotus japonicus*, *Glycine max*, *Medicago trunculata* got underway (Udvardi, 2002) for better understanding of the symbiotic interactions. Sequencing projects on these legumes focused on the large number of EST's (Expressed sequence tags). Udvardi (2002) reported that until 2002, the public database GenBank housed 255,291, 162,741 and 31,670 ESTs for these species. He further stated it has become a gold mine since EST collections symbolized cDNA libraries resulting either from specific organs obtained in different developmental stages, or from plants subjected to biotic and abiotic stress conditions and can be used for studying expression levels. The protein patterns in the diazotrophs are being studied using advanced proteomics approaches. Proteomic analyses have also revealed the direct genome functionality in a number of diazotrophs genomes (MacLean *et al.*, 2007). The understanding of the expression level of the genes is going on and is expected to throw light on the adaptation of the diazotrophs in soil or in symbiotic association. The completion of the genome projects for a large number of marine cyanobacteria as well as free-living diazotrophs like *Chlorobium, Clostridium, Azotobacter* and some methanogenic archaea has revealed new surprises and functional genomics approaches are being applied in a large scale. Studies concerning the nitrogen fixation in open oceans offer wealth of information about marine diazotrophs. There has been a revolution in the field of ocean genomics. The sequencing projects for *Synechococcus, Trichosdesmium* etc., discovery of novel genes yielded considerable information regarding their genome architecture in their respective habitats (Zehr *et al.*, 2000).

The research on functional genomics and proteomics for the nitrogen fixing bacteria is of great significance in the 21st century. It has become necessary to focus much on the comparative codon usage patterns, whole proteome analysis and molecular phylogeny using bioinformatics tools. Comparative genomics particularly focusing codon usage using different parameters is expected to provide insight into the inherent molecular nature of the genomes of diazotrophs.

WORKS ON CODON USAGE PATTERNS

In the post genomic era increasing number of genomes put forth a concept among computational biologists that each and every genome has its own story.

Since the time when the first nucleic acid sequences were obtained a number of hypothesis on the evolution of genomes have been put forward. The genetic code has been one of the most interesting aspects of biological science. The code is degenerate with multiple codons coding for a particular amino acid. Groups of codons coding for a particular amino acid are synonymous ones. It has been reported that these synonymous codons are somewhat conserved across species (Peden, 1999). The increase in sequence information *albeit* partial in the 1980's facilitated the studies concerning the usage of synonymous codons of organisms. Majority of work on codon usage patterns at that period focused upon *E. coli* (Peden, 1999). Gradually the techniques for codon usage were applied upon mammalian, bacterial, bacteriophage, viral and mitochondrial genes (Grantham *et al.*, 1980a, b, 1981). On the basis of studies conducted on mRNAs from a number of prokaryotic and eukaryotic species, Grantham *et al.* (1980a) proposed the "Genome Hypothesis" which hypothesized that codon usage pattern of a particular genome was an explicit attribute of that organism. Grantham *et al.* (1981) reported that difference in codon usage pattern might be associated with the tRNA content. More work on the codon usage patterns (Gouy and Gautier, 1982) in *E. coli* regarding codon usage and tRNA abundance led to the conclusion that highly expressed genes exhibited non-random codon usage and used a small set of codons that corresponded to abundant tRNAs. However, it was not clear why specific synonymous codons were used preferably. Grosjean and Fiers (1982) opined that optimal codon choice is the outcome of the necessity imposed by interaction between codon and cognate tRNA. Ikemura (1981) defined optimal codon as one that was translated by the most abundant cognate tRNA which he later amended (Ikemura, 1985). These optimal codons are under the influence of translational efficiency. Kurland (1991) reported that translational efficiency is shaped by highest turnover of ribosomes, effectiveness of aminoacyl-tRNA harmonizing and ternary complex conditions. Sharp *et al.* (1993) reported that the preference of some synonymous codons were the outcome of translational selection, i.e., for increasing efficiency and accuracy a codon is used that is translated by the abundant tRNA species. Rocha (2004) correlated codon usage bias from the tRNA point of view. He proposed that co-evolution of tRNA gene composition and codon bias in genomes from tRNA's point of view concur with the selection-mutation-drift theory. A number of studies (Sharp *et al.*, 1993; Carbone *et al.*, 2005) revealed that codon bias is influenced by effective population size, translational selection, mutational pressure, compositional bias and genetic drift. Chen *et al.* (2004) postulated that codon bias

is first and foremost influenced by mutational pressure and then translational selection. Studies on translational selection helped in the detection of highly expressed genes in genomes (Karlin and Mrazek, 2000). Majority of the highly expressed genes are associated with cell growth and cell division. It has been found that evaluating translational selection in quite difficult in genomes with a high or low G+C content owing to the consequence of intense levels of G+C content on codon usage. Medigue *et al.* (1991) applied the principles of codon usage, cluster analysis and correspondence analysis to study horizontal gene transfer mechanisms. A number of workers have suggested that codon usage and amino acid usage (Goldman and Yang, 1994; Nesti *et al.*, 1995; Schmidt, 1995) helps in determining the phylogenetic relationships between organisms. Besides, codon usage amino acid usage has also been studied at large (Peden, 1999). Ikemura (1981) was instrumental in reporting good correlation between amino acid composition and codon bias. Hydrophobicity, aromaticity and amino acid charges are testified to be influencing amino acid usage (Lobry and Gautier, 1994). The prediction of open reading frames has been performed utilizing the information of codon usage (Krogh *et al.*, 1994; Borodovsky *et al.*, 1995). Peden (1999) provided some classic examples of the utilization of GeneMark prediction programme (Borodovsky *et al.*, 1994) for identification of coding sequences from shotgun genome sequencing projects.

Sharp and Li (1987) were the pioneers in developing the Codon Adaptation Index (CAI) to assess the similarity amid the synonymous codon usage of a gene to that of the reference set. This particular index has been commonly used as a parameter for predicting the highly expressed genes in an organism. Besides CAI, a number of indices are regularly used to investigate codon usage patterns in a number of organisms. Peden (1999) opined that two types of indices one evaluating on the whole deviation of codon usage from the expected set and the other determining bias towards a specific subset of optimal codons are used. Indices such as GC content, GC3 content (Peden, 1999) effective number of codons (Nc) (Wright, 1990), relative synonymous codon usage (RSCU) (Sharp *et al.*, 1986) scaled Chi-square and G statistic, P2 index (Gouy and Gautier, 1982) measuring the codon-anticodon interaction, Codon Bias Index (CBI), Fop (frequency of optimal codons) (Ikemura, 1985) and codon adaptation index are very significant in studies concerning codon usage patterns. Besides these indices, correspondence analysis a type of multivariate statistical analysis are commonly used to find out the degree of associations between different genes and amino acids. It is a technique that explores non random synonymous codon usage.

This important field of study starved for good software in the 1980's. The basic problem was the integration of different indices into one program that would serve the researchers well. The foremost software was CODONS (Lloyd and Sharp, 1992). The GCG package (GCG, 1994) consisted of programs for analysis codon usage. CORRESPOND was used for performing multivariate statistical analysis (Peden, 1999). NetMul (Thioulouse *et al.*, 1995) was designed as a subset of ADE software for studying of codon usage and multivariate statistical analysis. CODON W (Peden, 1999) took over the stage and became very popular because of its robust nature and error free analysis. INCA (Supek and Vlahovicek, 2004) is also being used by researchers. Software was also developed for codon usage optimization. This involved alteration of rare codons in target genes so as to imitate the codon usage of host with no modification of the amino acid sequence of encoded proteins (Gustafsson *et al.*, 2004). GeneDesign (Richardson *et al.*, 2006), Synthetic Gene Designer (Wu *et al.*, 2006), Gene Designer (Villalobos *et al.*, 2006) are some of the software providing a platform for synthetic gene design and codon optimization. For estimation of the highly expressed genes using CAI there has been a number of software. CODON W (Peden, 1999), JCAT (Grote *et al.*, 2005), CAI CALCULATOR (Wu *et al.*, 2005), ACUA (Umashankar *et al.*, 2007) and e CAI server (Puigbo *et al.*, 2008) are commonly used. However, the e CAI server (Puigbo *et al.*, 2008) has proved to be the most powerful and efficient tool for estimation of expression levels of the genes.

Although a lot of work has been performed and is going on in codon usage of prokaryotes work on nitrogen fixing microorganisms has been lacking. Very preliminary work on codon usage of nitrogen fixing diazotrophs was initiated by Mather and Tuli (1991). McEwan and Gatherer (1999) used certain indices to predict gene functionality in an operon of *Frankia*. Ramseier and Gottfert (1991) reported differences in codon usage and GC content in *Bradyrhizobium* genes. Musto *et al.* (2003) accounted the role of translational selection for synonymous codon usage in *Clostridium acetobutylicum*, however they did not deal with the nitrogen fixation related genes. The authors' group has studied the synonymous codon usage patterns in *Frankia* EuIK1, *Bradyrhizobium japonicum* USDA, methanogenic archaea, three *Frankia* genomes, *Azotobacter vinelandii* and *Chlorobium* strains. Moderate codon bias was attributed to translational selection in nitrogen fixing genes of *Bradyrhizobium japonicum* USDA 110 (Sur *et al.*, 2005). Translational selection and mutational pressure influenced the high bias of *nif* genes for *Frankia* EuIK1. Correspondence analysis of RSCU values for *nif* genes indicated the role of GC3

compositional bias in effecting codon usage variation (Sur *et al.*, 2006). In case of the methanogenic archaea, the codon usage patterns revealed that nitrogen fixing genes were lowly biased and had low expression levels. Translational selection has been the source of variation of codon usage among nitrogen fixing genes. The author's group proposed that in methanogens *nif* genes did not undergo purifying selection and coevolved as a unit (Sur *et al.*, 2007). The analysis of synonymous codon usage patterns of three *Frankia* genomes (strains CcI3, ACN14a and EAN1pec) revealed that codon usage was highly biased, but variations were noticed among the three strains (Sen *et al.*, 2008). Using codon adaptation index (CAI) highly expressed genes in *Frankia* were predicted. Investigation of the predicted highly expressed genes demonstrated that *Frankia* strain CcI3 had a different profile compared to others. The findings supported the hypothesis that CcI3 is becoming a symbiotic specialist while the other two remained as facultative symbionts. The author's postulated that incidence of larger numbers of highly expressed genes in particular COG groups for strains ACN14a and EAN1pec gave them a chance to survive outside the symbiosis in contrast to CcI3. Synonymous codon usage analysis in *Azotobacter vinelandii* divulged considerable amount of heterogeneity. Highly expressed genes were rich in GC codons. About 503 potentially highly expressed genes were identified and most of them were linked to metabolic functions of which 10 were associated with the core nitrogen fixing mechanism (Sur *et al.*, 2008a,b). Codon usage by *Chlorobium* was moderately biased. Correlation of codon usage bias with tRNA content in *Chlorobium* genomes divulged the failure of the translational machinery to co-evolve with higher codon usage ensuing in moderate bias. Whole genome alignment using MAUVE depicted the conserved nature of the genomes. A huge number of potentially highly expressed genes in the *Chlorobium* strains were related to metabolism. Some genes like hydrogenases, nitrogenase iron protein complexes, bacteriochlorophylls and chlorosomes were highly expressed thus highlighting the survival patterns of the bacteria under strained environments (Sur *et al.*, 2008a,b).

PROTEOMICS AND STRUCTURAL BIOINFORMATICS ON NITROGEN FIXATION

High throughput sequencing of a number of eukaryotes and a number of bacteria has developed the science of proteomics and structural bioinformatics. Structural bioinformatics is expected to give rise to a large number of protein structures. The technological benefits of this field are already yielding results and have a substantial impact on structural biology research for prokaryotes (Burley and Bonnano, 2002). Rapid developments of technological know how in proteomics coupled with the improvement of bioinformatics tools have resulted in a deluge of structural information that guarantees acceleration in nitrogen fixation research.

Proteomics techniques have been used to explore legume-microbe interactions. The proteomic studies are focusing on determination of the influence of microorganisms on host roots, for identifying genes and proteins induced during root-microbe interactions (Mathesius, 2009). Most of the research has dealt with legume-symbioses for nitrogen-fixing rhizobia. It has been possible to detect specific microbial signal molecules in plants, harmonizing of signal responses and modification of plant growth by microorganisms. Combinations of 2D gel, peptide mass fingerprinting and bioinformatics aided in understanding of factors linked to interactions of *Rhizobium, Mesorhizobium, Sinorhizobium* with model legume plants like *Medicago trunculata* and *Lotus japonicus* (Rolfe *et al.*, 2003). Investigation concerning acid tolerance of *Bradyrhizobium japonicum* USDA 110 using 2-D gel electrophoresis and MALDI-TOF techniques offered information regarding protein compositions under variable pH and illustrated that acid tolerant property of *Bradyrhizobium japonicum* had relationships with a number of important metabolic pathways (Puranamaneewiwat *et al.*, 2006). A number of proteomic approaches were used to identify cyanobacterial symbiotic associations in *Azolla*. The analysis revealed some metabolisms that were committed to nitrogen fixation and expression of some vital proteins earlier unknown to influence cyanobacterial symbiosis. Further analysis exposed the nature of modifications associated with Nif H proteins in the cyanobacterial (Ekman *et al.*, 2008). Proteomics studies of *Gluconacetobacter diazotrophicus* in exponential and stationary phases of cultures in presence of low and high levels of combined nitrogen revealed 131 differentially expressed proteins (Lery *et al.*, 2008). Proteins associated with cell wall biogenesis were reported (Lery *et al.*, 2008) to be expressed under limited nitrogen and shaped functioning of nitrogen-fixing cells. Nitrogen fixation accessory proteins were produced when cells grew under limited nitrogen and these proteins helped in stabilization of nitrogenase as well its protection from oxygen. Comparative proteomic analysis of *Rhodospirillum rubrum* identified 44 differentially expressed proteins. Proteins stimulated in nitrogen-fixing conditions are associated with nitrogen fixation and assimilation and are consistent with Nif regulated expressions (Selao *et al.*, 2008).

Although a number of works has been performed using X-ray crystallography and NMR spectroscopy for *Azotobacter* and *Clostridium* nitrogenases (Burgess and Lowe, 1996; Rees and Howard, 2000; Lawson and Smith, 2002; Seefeldt *et al.*, 2004) there are absence of comprehensive studies using homology modeling techniques. This might have been due to the non-availability of appropriate software and tools for validations. The difficulties in ascertaining the three-dimensional structures of proteins by spectroscopy and X-ray crystallographic analysis have made computational techniques very popular (Othman *et al.*, 2007). Homology modelling has become a dependable procedure for consistently predicting 3D structure of proteins with accuracy levels akin to those attained at low-resolution using experimental methods (Marti-Renom *et al.*, 2000). The method aligns the amino acid sequence of unknown structure (target) with homologue of known structure (template). In view of the fact, that study of 3D structure of a protein is supportive in making out the finer points of a protein (Paramsivasan *et al.*, 2006), this method has become wide spread in bioinformatics. Homology models of Nif H proteins from *Trichodesmium* sp. IMS101 have reported by Zehr *et al.* (1997a) and the implications of aerobic nitrogen fixation were analyzed. Sen and Peters (2005) reported the homology model of the Nif H protein from *Methanobacter thermoautotrophicus*. The structure demonstrated the interplay of a number of factors playing a role in the thermal adaptation of protein-protein interactions in *M. thermoautotrophicus*. The structure has been reported to be rigid and assisted in thermal stability.

The authors' group has been the pioneer in determination of the 3D structure of *Frankia* Nif H proteins in three strains viz., *Frankia* sp. CcI3, EAN1pec and ACN14a using homology modeling technique (Sen *et al.*, 2010). Metal binding sites and functionally imperative section of the proteins were investigated. Thiol ligands and active sites helped in protein functioning and conformations. Structurally essential nests were predicted. The clefts and cavities retained biologically important residues. *In silico* site-directed mutagenesis analyses divulged that mutations in functional residues hamper nitrogen fixation. The structures were rigid with accessible area for solvent interactions. The models are consistent offering insights into the 3D structural framework as well as structure-function relationships. However, most of the studies on the protein complement of the sequenced diazotrophs have largely focused on functional annotations and amino acid compositions with little focus on common physical properties. The steady increase in complete proteomes for diazotrophs pose a

significant challenge for accomplishing study in large numbers. Thus, a broad picture on protein signatures for nitrogen fixing microorganisms is unavailable. The physical properties of proteins are vital for functioning of an organism (Knight *et al.*, 2004). An understanding of how these properties render the role of complete proteomes is expected to fetch answers relevant to organism and molecular biology (Knight *et al.*, 2004).

The author's group has studied the physical properties of proteins like isoelectric point, protein energetic cost, aromaticity and hydropathicity and amino-acid adaptation index in portraying the nature of proteomes and consequently lifestyle of diazotrophs (unpublished). The recognized diversity in amino-acid usage patterns, the forces influencing them, investigated the association of isoelectric point with lifestyle of the organisms shaping its role in manipulating diversity across proteomes. Interrelationships between isoelectric point, protein energetic cost, aromaticity, hydropathicity and amino acid adaptation index were noted. Amino acid usage is influenced by hydropathicity and aromaticity. Distribution of isoelectric point across COGs functional groups sheds light upon diversity and lifestyle of the organisms and establishes isoelectric point as a molecular signature. Nitrogen fixing proteomes are influenced by their environmental adaptations. COG functional groups belonging to cellular processes, metabolism and information storage processing shape acidic proteomes of nitrogen fixers.

PHYLOGENETIC STUDIES ON BNF

Nothing in biology makes sense without evolution. This also applies for biological nitrogen fixation. Earlier studies of biological evolution mainly concentrated either upon general appearance or habitats or genetic similarities (Williams, 2007). The foundation and current distribution of nitrogen fixation has been mystifying from a phylogenetic point of view owing to a number of factors that puzzled the scientists for decades (Raymond *et al.*, 2004). Raymond *et al.* (2004) reported that the ancient phenomenon of nitrogen fixation played a vital role in the spreading out of microbial life on our planet. All nitrogen fixing diazotrophs have a wide distribution in the kingdoms of Bacteria and Archaea. A number of workers (Raven and Yin, 1998; Kasting and Siefert, 2001; Navarro-Gonzalez *et al.*, 2001) hypothesized that abiotic nitrogen was a limiting factor in the early days of the Earth and reduced availability of nitrogen influenced the evolution of early nitrogen fixers. The complex climatic, geological and biochemical events of the early atmosphere too played a significant role in the evolution of nitrogen fixers.

In the post genomic era, the accessibility of the genes, proteins and completely sequenced genomes have made it feasible to comprehend the impact of complex genetic events like paralogy, lateral gene transfer, gene duplication events in the evolution of nitrogen fixation. It has now become possible to understand clearly the forces influencing the evolution of nitrogenase at the molecular level. Raymond *et al.* (2004) reported that nitrogenase evolved in multiple lineages and there are evidences of loss, duplications and horizontal and vertical transfers for the nitrogenase genes and operons during the course of evolution. Till date most of the studies concerning the evolution of nitrogen fixation have focused on the *nif* genes since they are extensive, found in cotranscribed regulons, house the major subunits for the nitrogenase and code for a number of proteins involved in the regulation and activation of the nitrogen fixing mechanism (Kessler *et al.*, 1998; Halbleib and Ludden, 2000; Goodman and Weisz, 2002). Owing to the far-reaching efforts of scientists throughout the world it has been possible to sequence the nitrogenase genes like *nif* H, *nif* D, *nif* K, *nif* E, *nif* N etc. for understanding the evolutionary history of nitrogen fixation (Normand and Bousquet, 1989; Normand *et al.*, 1992; Hirsch *et al.*, 1995; Zehr *et al.*, 1997b; Fani *et al.*, 2000). There has been a rigorous debate on the evolution of these genes among different workers (Normand and Bousquet, 1989; Fani *et al.*, 2000; Zehr *et al.*, 2003; Henson *et al.*, 2004; Kechris *et al.*, 2006). These debates focused on issues ranging from random distribution of nitrogen fixation (Henson *et al.*, 2004) to lateral gene transfer (Normand and Bousquet, 1989; Kechris *et al.*, 2006), vertical descent accompanied by multiple losses (Young, 1992; Normand *et al.*, 1992; Hirsch *et al.*, 1995), convergent evolution (Postgate and Eady, 1988) etc. There have been conflicting reports on the evolution of the genes arising out of investigations with dissimilar techniques of phylogenetic analysis (Henson *et al.*, 2004). Some workers (Normand and Bousquet, 1989; Normand *et al.*, 1992; Hirsch *et al.*, 1995; Hennecke *et al.*, 1992; Zehr *et al.*, 1998, 2000) reported contradictory information on the agreement between *nif* H phylogenies and 16S rRNA based phylogenies in cyanobacteria, proteobacteria and actinobacteria while others (Henson *et al.*, 2004) investigated the evolutionary history of nitrogen fixation using only *nif* D. Most of the workers used the sequence alignment based methods to study evolution of relevant *nif* genes.

Normand and Bousquet (1989) were the first to hypothesize that nitrogen fixation arose once in the course of evolution and has been transmitted horizontally to different lineages. The other hypothesis that cropped up in the last decade of the 20th century, predicted that nitrogen fixation arose in the early evolutionary history and was omnipresent in all prokaryotic organisms at that time. Gradually with time it disappeared in some lineages and presently retained in some (Young, 1992; Normand *et al.*, 1992; Hirsch *et al.*, 1995). Fani *et al.* (2000) postulated a model that in some ways recreated the evolutionary history of *nif* D, *nif* K, *nif* E and *nif* N. According Fani *et al.* (2000), the primitive ancestral gene arose by means of paralogous duplication events. This was followed by deviation to an ancestral bicistronic operon and this operon further went through paralogous operon duplication events thus creating the modern ancestors of *nif* genes. They further hypothesized that *nif DKEN* genes were already present in the earliest prokaryotes prior to the division of archaea and bacteria and weakening of selection pressure owing to higher energetic costs of the process of nitrogen fixation resulted in loss of *nif* genes in the late prokaryotes thus limiting them to a few. Henson *et al.* (2004) re-examined the phylogeny of nitrogen fixation by analyzing only the molybdenum containing *nif* D gene from a cyanobacteria, proteobacteria as well as gram-positive bacteria. They compared the *nif* D phylogeny with that of the 16S rRNA phylogeny. The rationale behind this was that in cases where *nif* and rRNA topologies matched, vertical descent could be inferred otherwise lateral gene transfer could be the reason. They applied the parsimony technique, maximum likelihood technique and neighbor joining technique to analyze the evolution of the *nif* D genes. Henson *et al.* (2004) inferred that their studies on *nif* D phylogenies are harmonious with that of the 16S rRNA based phylogenies and supported the hypothesis of vertical descent. Their results also matched with the *nif* H phylogenies (Zehr *et al.*, 1998, 2000) with respect to vertical descent. Henson *et al.* (2004) reported that parsimony analysis of the amino acid sequences and maximum likelihood studies concerning the nucleotide sequences of *nif* D genes support monophyletic origins of cyanobacterial and actinobacterial strains strengthening their hypothesis on vertical descent. Conversely, distance matrix analysis of nucleotide sequences for *nif* D genes bolstered the concept of horizontal gene transfer. They also argued that the parsimony and maximum likelihood methodologies are probably more superior compared to other methodologies based on sequence alignments for resolving the evolutionary relationships between organisms. Kechris *et al.* (2006) however strongly lent support for horizontal gene transfer events.

The authors' group have a skeptical view with respect to sequence alignment based methods since

alignment based methods and structure based techniques are not fully adequate to throw light on the complex evolutionary process of nitrogen fixation. Phylogenetic techniques based on sequence alignment and structures are inadequate in studying evolution since, sequence similarity becomes erratic at identity levels below 25% (Mondol *et al.*, 2008). Some workers (Qi *et al.*, 2004; Sims *et al.*, 2008) have also pointed out fallacies in sequence alignment based methods. Again, alignment based methods fail to infer relationships from whole genome data (Qi *et al.*, 2004). Structure based techniques have incomplete span given that number of structures are inadequate to portray a significant conclusion. The author's group has devised an alignment-free, condensed matrix method relying on nucleotide triplet based phylogeny (Randic *et al.*, 2001; Randic, 2000; Mondal *et al.*, 2008; Sur *et al.*, 2009) to re-examine the evolution of nitrogen fixation (unpublished). They have considered all three codon positions in nucleotide triplets, giving them equal weight, thus incorporating their influence during evolution. This novel technique takes into account a set of invariants in a DNA sequence and determines the extent of resemblance among DNA sequences using the invariants (Randic *et al.*, 2001). The frequency of incidence of bordering base pairs (Randic, 2000) is calculated by DNA portrayed by condensed matrices. These matrices infer upon the characteristics of DNA primary sequence and recognize qualitative and quantitative differences between DNA sequences at the intraspecific and interspecific level (Randic, 2000). The nucleotide sequences of the genes were examined to find out the relationship between them. The matrices linked to each sequence were determined and the leading eigen values calculated. The leading eigenvectors help in analyzing DNA sequence for resemblance and contrast and the correlation between leading eigen values were analyzed by variations between them (Nandy *et al.*, 2006). Re-investigation of the evolutionary history of nitrogen fixation was done by analyzing *nif* H, D and K genes from cyanobacteria, proteobacteria, gram positive bacteria, green sulfur bacteria and archaea. Evaluation of *nif* H, D and K phylogenies with whole genome phylogenies and 16S rRNA phylogeny was carried out using the same method since, rRNA based phylogenies are assumed to be precise organismal phylogeny (Henson *et al.*, 2004). The findings indicated that structural properties of DNA sequences are guided by different descriptors and invariants. Results supported evidence for polyphyletic origin, horizontal gene transfer and gene duplication events as potent evolutionary forces. Duplications in portions of genes, operons or stretches of nucleotide sequences might have had occurred during transformation of primitive nitrogenase to the present form, playing a crucial role in producing genes with similar properties. There is agreement on *nif* H, D and K phylogenies, with respect to horizontal gene transfer and gene duplication events. This work throws light on better understanding of the diversity of nitrogen fixation and the technique employs the power of categorization of DNA sequences by invariants to recognize the qualitative and quantitative properties.

FUTURE PROSPECTS

Nitrogen fixation research presents a completely new outlook in the post genomic era. The wealth of information obtained from the genome projects needs to be mined. As newer and newer diazotrophs are discovered and sequenced novel insights are being gained. Bioinformatics in combination with metagenomics as well as metaproteomics approaches has the potential to give a more detailed picture that underlie symbiotic as well as asymbiotic nitrogen fixation. At the end of the day scientific perception will continue to play a vital role in creating models that clarify nitrogen fixation in improved manner.

ACKNOWLEDGMENT

REFERENCES

Alexander, M., 1977. Introduction to soil microbiology, 2nd Edn., John Wiley and Sons, New York..

Arnold, W., A. Rump, W. Klipp, U.B. Priefer and A. Pühler, 1988. Nucleotide sequence of a 24,206-base-pair DNA fragment carrying the entire nitrogen fixation gene cluster of *Klebsiella pneumonia*. J. Mol. Biol., 203: 715-738.

Baladani, I.J. and L.V. Baldani, 2005. History on the biological nitrogen fixation research in graminaceous plants: Special emphasis on the Brazilian experience. Anais Da Acad. Brasiliera De Ciencias Sci., 77: 549-579.

Belay, N., R. Sparling and L. Daniels, 1984. Dinitrogen fixation by a thermophilic methanogenic bacterium. Nature, 312: 286-288.

Benedito, V.A., X. Dai, J. He, X.P. Zhao and M.K. Udvardi, 2006. Functional genomics of plant transporters in legume nodules. Func. Plant Biol., 33: 731-736.

Benson, D.R. and W.B. Silvester, 1993. Biology of *Frankia* strains, actinomycete symbionts of actinorhizal plants. Microbiol. Rev., 57: 293-319.

Birney, E., N. Clamp and T. Hubbard, 2002. Databases and tools for browsing genomes. Ann. Rev. Genomics Hum. Genet., 3: 293-310.

Bloom, M., 2001. Biology *in silico*: The bioinformatics revolution. Am. Biol. Teacher, 63: 397-403.

Bohlool, B.B., J.K. Ladha, D.P. Garrity and T. George, 1992. Biological nitrogen fixation for sustainable agriculture. Plant Soil, 141: 1-11.

Borodovsky, M., K.E. Rudd and E.V. Koonin, 1994. Intrinsic and extrinsic approaches for detecting genes in a bacterial genome. Nucleic Acids Res., 22: 4756-4767.

Borodovsky, M., J.D. McIninch, E.V. Koonin, K.E. Rudd, C. Medigue and A. Danchin, 1995. Detection of new genes in a bacterial genome using Markov models for 3 gene classes. Nucleic Acids Res., 23: 3554-3562.

Bumb, B. and C.A. Baanante, 1996. The Role of Fertilizer in Sustaining Food Security and Protecting the Environment to 2020. International Food Policy Research Institute, Washington DC.

Burgess, B.K. and D.J. Lowe, 1996. Mechanism of molybdenum nitrogenase. Chem. Rev., 96: 2983-3011.

Burley, S.K. and J.B. Bonnano, 2002. Structuring the universe of proteins. Ann. Rev. Genomics Hum. Genet., 3: 243-262.

Callaham, D., P. Del Tredici and J.G. Torrey, 1978. Isolation and cultivation *in vitro* of the actinomycetes causing root nodulation in *Comptonia*. Science, 199: 899-902.

Carbone, A., F. Kepes and A. Zinovyev, 2005. Codon bias signatures, organization of microorganisms in codon space and lifestyle. Mol. Biol. Evol., 22: 547-561.

Chen, S.L., W. Lee, A.K. Hottes, L. Shapiro and H.H. McAdams, 2004. Codon usage between genomes is constrained by genome-wide mutational processes. Proc. Nat. Acad. Sci. USA., 101: 3480-3485.

Costa, J.L. and P. Lindblad, 2002. Cyanobacteria in Symbiosis in Cycads. In: Cyanobacteria in Symbiosis, Rai, A.N., B. Bergman and U. Rasmussen, (Eds.). Kluwer Academic Publ., Dordrecht, pp: 195-206.

Cérémonie, H., B. Cournoyer, F. Maillet, P. Normand and M.P. Fernandez, 1998. Genetic complementation of rhizobial nod mutants with *Frankia* DNA: Artifact or realty?. Mol. Gen. Genet., 260: 115-119.

Dixon, R. and D. Kahn, 2004. Genetic regulation of biological nitrogen fixation. Nat. Rev. Microbiol., 2: 621-631.

Dobereiner, J., 1992. History and new perspectives of diazotrophs in association with nonleguminous plants. Symbiosis, 13: 1-13.

Einsle, O., F.A. Tezcan, S.L.A. Andrade, B. Schmid, M. Yoshida, J.B. Howard and D.C. Rees, 2002. Nitrogenase MoFe protein at 1.16 angstrom resolution: A central ligand in the FeMo-Cofactor. Science, 297: 1696-1700.

Ekman, M., P. Tollback and B. Bergman, 2008. Proteomic analysis of the cyanobacterium of the *Azolla* symbiosis: Identity, adaptation and Nif H modification. J. Exp. Bot., 59: 1023-1034.

Elmerich, C., 1984. Molecular biology and ecology of diazotrophs associated with non-leguminous plants. Nat. Biotechnol., 2: 967-978.

Fani, R., R. Gallo and P. Lio, 2000. Molecular evolution of nitrogen fixation: The evolutionary history of the nifD, nifK, nifE and nifN genes. J. Mol. Evol., 51: 1-11.

Fleischmann, R.D., M.D. Adams, O. White, R.A. Clayton and E.F. Kirkness *et al.*, 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science, 269: 496-512.

Franche, C., K. Lindstrom and C. Elmerich, 2009. Nitrogen fixing bacteria associated with leguminous and non-leguminous plants. Plant Soil, 321: 35-59.

Fujita, K., K.G. Otsu and S. Ogta, 1992. Biological nitrogen fixation in mixed legume-cereal cropping systems. Plant Soil, 141: 155-175.

GCG, 1994. Programme Manual for the Wisconsin Package. Version 8, Genetics Computer Group, Wisconsin, USA.

Georgiadis, M.M., H. Komiya, P. Chakraborti, D. Woo, J.J. Kornuc and D.C. Rees, 1992. Crystallographic structure of the nitrogenase iron protein from *Azotobacter vinelandii*. Science, 257: 1653-1659.

Geurts, R. and T. Bisseling, 2002. Rhizobium nod factor perception and signaling. Plant Cell, 14: 239-249.

Goldman, N. and Z.H. Yang, 1994. Codon based model of nucleotide substitution for protein coding DNA sequences. Mol. Biol. Evol., 11: 725-736.

Gonzalez, V., R.I. Santamaria, P. Bustos, I. Hernandez-Gonzalez, A. Medrano-Soto and G. Moreno-Hagelsieb *et al.*, 2006. The partitioned *Rhizobium etli* genome: genetic and metabolic redundancy in seven interacting replicons. Proc. Natl. Acad. Sci. USA., 103: 3834-3839.

Goodman, R.M. and J.B. Weisz, 2002. Plant-Microbe Symbioses: An Evolutionary Survey. In: Biodiversity of Microbial Life: Foundation of Earths Biosphere, Staley, J.T. and A.L. Reysenbach (Eds.). Wiley, New York.

Gouy, M. and C. Gautier, 1982. Codon usage in bacteria correlation with gene expressivity. Nucleic Acids Res., 10: 7055-7074.

Grantham, R., C. Gautier and M. Gouy, 1980a. Codon frequencies in 119 genes confirm consistent choices of degenerate base according to genome type. Nucleic Acids Res., 8: 1893-1912.

Grantham, R., C. Gautier, M. Gouy, R. Mercier and A. Pave, 1980b. Codon catalogue usage and the genome hypothesis. Nucleic Acids Res., 8: r49-r62.

Grantham, R., C. Gautier, M. Gouy, M. Jacobzone and R. Mercier, 1981. Codon catalogue usage is a genome strategy for genome expressivity. Nucleic Acids Res., 9: 213-213.

Gresshoff, P.M., 2003. Post genomic insights into plant nodulation symbiosis. Genome Biol., 4: 201-201.

Grosjean, H. and W. Fliers, 1982. Preferential codon usage in prokaryotic genes: The optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes. Gene, 18: 209-299.

Grote, A., K. Hiller, M. Scheer, R. Munch, B. Nortemann, D.C. Hempel and D. Jahn, 2005. JCat: A novel tool to adapt codon usage of a target gene to its potential expression host. Nucleic Acids Res., 33: W526-W531.

Gubler, M. and H. Hennecke, 1986. Fix A, B and C are essential for symbiotic and free living microaerobic nitrogen fixation. FEBS Lett., 200: 186-192.

Gustafsson, C., S. Govindarajan and J. Minshull, 2004. Codon bias and heterologous protein expression. Trends Biotechnol., 22: 346-353.

Halbleib, C.M. and P.W. Ludden, 2000. Regulation of biological nitrogen fixation. J. Nutr., 130:1081–1084.

Hamelin, J., N. Fromin, S. Tarnawski, S. Teyssier-Cuvelle and M. Aragno, 2002. nifH gene diversity in the bacterial community associated with the rhizosphere of *Molinia coerulea*, an oligonitrophilic perennial grass. Environ. Microbiol., 4: 477-481.

Hennecke, H., K. Kaluzu, B. Thony, M. Fuhrmann, W. Ludwig and E. Stackebrandt, 1992. Concurrent evolution of nitrogenase genes and 16S rRNA in *Rhizobium* species and other nitrogen fixing bacteria. Arch. Microbiol., 142: 342-348.

Henson, B.J., L.E. Watson and S.R. Barnum, 2004. The evolutionary history of nitrogen fixation, as assessed by nifD. J. Mol. Evol., 58: 390-399.

Hirsch, A.M., H.I. McKhann, A. Reddy, J. Liao, Y. Fang and C.R. Marshall, 1995. Assessing horizontal transfer of analysis using quartets and parallel computing. Bioinformatics, 18: 502-504.

Howard, J.B. and D.C. Rees, 1996. Structural basis of biological nitrogen fixation. Chem. Rev., 96: 2965-2982.

Hu, Y., A.W. Fay, C.C. Lee and M.W. Ribbe, 2007. P-cluster maturation on nitrogenase MoFe protein. Proc. Natl. Acad. Sci. USA., 104: 10424-10429.

Ideker, T., T. Galitski and L. Hood, 2001. A new approach to decoding life: Systems biology. Ann. Rev. Genom. Hum. Genet., 2: 343-372.

Ikemura, T., 1981. Correlation between the abundance of *Escherichia coli* transfer. RNAs and the occurrence of the respective codons in its protein genes. J. Mol. Biol., 146: 1-21.

Ikemura, T., 1985. Codon usage and transfer-RNA content in unicellular and multicellular organisms. Mol. Biol. Evol., 2: 13-34.

Kaneko, T., Y. Nakamura, S. Sato, K. Minamisawa and T. Uchiumi *et al.*, 2002. Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110. DNA Res., 9: 189-197.

Karlin, S. and J. Mrazek, 2000. Predicted highly expressed genes of diverse prokaryotic genomes. J. Bacteriol., 182: 5238-5250.

Kasting, J.F. and J.L. Siefert, 2001. Biogeochemistry: The nitrogen fix. Nature, 412: 26-27.

Kechris, K.J., J.C. Lin, P.J. Bickel and A.N. Glazer, 2006. Quantitative exploration of lateral gene transfer by using nitrogen fixation genes as a case study. Proc. Natl. Acad. Sci., 103: 9584-9589.

Kessler, P.S., C. Blank and J.A. Leigh, 1998. The nif gene operon of the methanogenic archaeon *Methanococcus maripaludis*. J. Bacteriol., 180: 1504-1511.

Keyser, H.H. and L. Fudi, 1992. Potential for increasing biological nitrogen fixation in soybean. Plant Soil, 141: 119-135.

Kim, J., D. Woo and D.C. Rees, 1993. X-ray crystal structure of the nitrogenase molybdenum- protein from *Clostridium pasteurianum* at 3.0 angstrom resolution. Biochemistry, 32: 7104-7115.

Knight, C.G., R. Kassen, H. Hebestreit and P.B. Rainey, 2004. Global analysis of predicted proteomes: Functional adaptation of physical properties. Proc. Nat. Acad. Sci. USA., 101: 8390-8395.

Krogh, A., I.S. Mian and D. Haussler, 1994. A hidden Markov model that finds genes in *Escherichia coli* DNA. Nucleic Acids Res., 22: 4768-4778.

Kurland, C.G., 1991. Codon bias and gene-expression. FEBS Lett., 285: 165-169.

Kyrpides, N.C., 1999. Genomes online database (GOLD 1.0): A monitor of complete and ongoing genome projects worldwide. Bioinformatics, 15: 773-774.

Ladha, J.K. and P.M. Reddy, 1995. Extension of nitrogen fixation to rice-necessity and possibilities. Geojournal, 35: 363-372.

Lawes, J.B. and J.H. Gilbert, 1895. The Rothamsted Experiments Over 50 years. Blackwood and Sons, Edinburgh, London.

Lawson, D.M. and B.E. Smith, 2002. Molybdenum nitrogenases: A crystallographic and mechanistic view. Met. Ions Biol. Syst., 39: 75-119.

Lechevalier, M.P., 1994. Taxonomy of the genus *Frankia* (Actinomycetales). Int. J. Syst. Bacteriol., 44: 1-8.

Ledgard, S.F. and K.W. Steele, 1992. Biological nitrogen fixation in mixed legume/grass pastures. Dev. Plant Soil Sci., 49: 137-153.

Lery, L.M.S., W.M.A. von Krüger, F.C. Viana, K.R.S. Teixelra and P.M. Bisch, 2008. A comparative proteomic analysis of *Gluconacetobacter diazotrophicus* PAL 5 at exponential and stationary phases cultures in the presence of high and low levels of inorganic nitrogen compound. Biochimi. Biophys. Acta, 1784: 1578-1589.

Lloyd, A.T. and P.M. Sharp, 1992. CODONS: A microcomputer program for codon usage analysis. J. Hered., 83: 239-240.

Lobry, J.R. and C. Gautier, 1994. Hydrophobicity, expressivity and aromaticity are the major trends of amino-acid usage in 999 *Escherichia coli* chromosome-encoded genes. Nucleic. Acids Res., 22: 3174-3180.

Long, S.R., 2001. Gene and signals in the *Rhizobium*-legume symbiosis. Plant Physiol., 125: 69-72.

MacLean, A.M., T.M. Finan and M. J. Sadowsky, 2007. Genomes of the symbiotic nitrogen-fixing of Bacteria legumes. Plant Physiol., 144: 615-622.

Marti-Renom, M.A., A.C. Stuart, A. Fiser, R. Sanchez, F. Melo and A. Sali, 2000. Comparative protein structure modeling of genes and genomes. Ann. Rev. Biophys. Biomol. Struct., 29: 291-325.

Mather, M. and R. Tuli, 1991. Analysis of codon usage in genes for nitrogen fixation from phylogenetically diverse diazotrophs. J. Mol. Evol., 32: 364-373.

Mathesius, U., 2009. Comparative proteomic studies of root-microbe interactions. J. Proteomics, 72: 353-366.

McEwan, N.R. and D. Gatherer, 1999. Codon indices as a predictor of gene functionality in a *Frankia* operon. Can. J. Bot., 77: 1287-1292.

Medigue, C., T. Rouxel, P. Vigier, A. Henaut and A. Danchin, 1991. Evidence for horizontal gene transfer in *Escherichia coli* speciation. J. Mol. Biol., 222: 851-856.

Merrick, M.J. and R.A. Edwards, 1995. Nitrogen control in bacteria. Microbiol. Rev., 59: 604-622.

Miflin, B.J. and P.J. Lea, 1977. The pathway of nitrogen assimilation in plants. Progr. Phytochem., 4: 1-26.

Miller, W., D.M. Kateryna, A. Nekrutenko and R.C. Hardison, 2004. Comparative genomics. Ann. Rev. Genomics Hum. Genet., 5: 15-56.

Mondol, U.K., B. Das, T.C. Ghosh, A. Sen and A.K. Bothra, 2008. Nucleotide triplet based molecular phylogeny of classI and classII aminoacyl t-RNA synthetase in three domain of life process: bacteria, archaea and eukarya. J. Biomol. Struct. Dyn., 26: 321-328.

Murray, P.A. and S.H. Zinder, 1984. Nitrogen fixation by a methanogenic archaebacterium. Nature, 312: 284-286.

Musto, H., H. Romero and A. Zavala, 2003. Translational selection is operative for synonymous codon usage in *Clostridium perfringens* and *Clostridium acetobutylicum*. Microbiology, 149: 855-863.

Mylona, P., K. Pawlowski and T. Bisseling, 1995. Symbiotic nitrogen fixation. Plant Cell, 7: 869-885.

Nandy, A., M. Harle and S.C. Basak, 2006. Mathematical descriptors of DNA sequences: development and applications ARKIVOC, ix: 211-238.

Navarro-Gonzalez, R., C.P. McKay and D.N. Mvondo, 2001. A possible nitrogen crisis for Archaean life due to reduced nitrogen fixation by lightning. Nature, 412: 61-64.

Nesti, C., G. Poli, M. Chicca, P. Ambrosino, C. Scapoli and I. Barrai, 1995. Phylogeny inferred from codon usage pattern in 31 organisms. Comput. Applied Biosci., 2: 167-171.

Newton, W.E., 2007. Physiology, Biochemistry and Molecular Biology of Nitrogen Fixation. In: Biology of the Nitrogen Cycle, Bothe, H., S.J. Ferguson and W.E. Newton (Eds.). Elsevier, Amsterdam, pp: 109-130.

Normand, P. and J. Bousquet, 1989. Phylogeny of nitrogenase sequences in *Frankia* and other nitrogen-fixing microorganisms. J. Mol. Evol., 29: 436-444.

Normand, P., M. Gouy, B. Cournoyer and P. Simonet, 1992. Nucleotide sequence of nif D from *Frankia alni* strain ArI3: Phylogenetic inferences. Mol. Biol. Evol., 9: 495-506.

Normand, P., P. Lapierre, L.S. Tisa, J.P. Gogarten and N. Alloisio *et al.*, 2007. Genome characteristics of facultatively symbiotic *Frankia* sp. strains reflect host range and host plant biogeography. Genome. Res., 17: 7-15.

Othman, R., H.A. Wahab, R. Yosof and N.A. Rahman, 2007. Analysis of secondary structure predictions of dengue virus type 2 NS2B/ NS3 against crystal structure to evaluate the predictive power of the *in silico* methods. *In Silico* Biol., 7: 215-224.

Paramsivasan, R., R. Sivaperumal, K.J. Dhnanjeyan, V. Thenmozhi and B.K. Tyagi, 2006. Prediction of 3-dimensional structure of salivary odorant-binding protein-2 of the mosquito *Culex quinquefasciatus*, the vector of human lymphatic filariasis. *In Silico* Biol., 7: 1-6.

Peden, J., 1999. Analysis of codon usage. Ph.D. Thesis. The University of Nottingham, United Kingdom.

Perez-Iratxeta, C., M.A. Andrade-Navarro and J. D. Wren, 2007. Evolving research trends in bioinformatics. Brief. Bioinform., 8: 88-95.

Perret, X., C. Staehelin and W.J. Broughton, 2000. Molecular basis of symbiotic promiscuity. Mol. Biol. Rev., 64: 180-201.

Peters, J.W. and R.K. Szilagyi, 2006. Exploring new frontiers of nitrogenase structure and mechanism. Curr. Opin. Chem. Biol., 10: 101-108.

Postgate, J.R. and R.R. Eady, 1988. Nitrogen fixation: Hundred years after. Proceedings of the 7th International Congress on Nitrogen Fixation, March 13-20, Gustav Fischer, Stuttgart, pp: 31-40.

Puhler, A., M. Arlat, A. Becker, M. Gottfert, J.P. Morrissey and F. O'Gara, 2004. What can bacterial genome research teach us about bacteria-plant interactions?. Curr. Opin. Plant. Biol., 7: 137-147.

Puigbo, P., I.G. Bravo and S. Vallve-Garcia, 2008. E-CAI: A novel server to estimate an expected value of codon adaptation index (e CAI). BMC Bioinformatics, 9: 65-65.

Puranamaneewiwat, N., S. Tajima and H. Niamsup, 2006. Proteomic analysis of *Bradyrhizobium japonicum* USDA 110 in acidic condition. Chiang Mai J. Sci., 33: 335-346.

Qi, J., B. Wang and B.I. Hao, 2004. Whole proteome prokaryote phylogeny without sequence alignment: A K-string composition approach. J. Mol. Evol., 58: 1-11.

Ramseier, T.M. and M. Gottfert, 1991. Codon usage and G+C content in *Bradyrhizobium japonicum* genes are not uniform. Arch. Microbiol., 156: 270-276.

Randic, M., 2000. Condensed representation of DNA primary sequences. J. Chem. Infrom. Comput. Sci., 40: 50-56.

Randic, M., X. Guo and S.C. Basak, 2001. On the characterization of DNA primary sequences by triplet of nucleic acid bases. J. Chem. Inform. Comput. Sci., 41: 619-626.

Rao, N.S.S., 1988. Ten decades of research on biological nitrogen fixation. Curr. Sci., 57: 710-713.

Raven, J.A. and Z.H. Yin, 1998. The past, present and future of nitrogenous compounds in the atmosphere and their interactions with plants. New Phytol., 139: 205-219.

Raymond, J., J.L. Siefert, C.R. Staples and R.E. Blankenship, 2004. The natural history of nitrogen fixation. Mol. Biol. Evol., 21: 541-554.

Rees, D.C. and J.B. Howard, 2000. Nitrogenase: Standing at the crossroads. Curr. Opin. Chem. Biol., 4: 559-566.

Richardson, S.M., S.J. Wheelan, R.M. Yarrington and J.D. Boeke, 2006. Gene design: Rapid, automated design of multi kilobase synthetic genes. Genome. Res., 16: 550-556.

Ritchie, G.A.F. and D.J.D. Nicholas, 1972. Identification of the sources of nitrous oxide produced by oxidation and reductive processes in *Nitrosomonas europaea*. Biochem. J., 126: 1181-1191.

Rocha, E.P.C., 2004. Codon usage bias from tRNAs point of view: Redundancy, specialization and efficient decoding for translation optimization. Genome Res., 14: 2279-2286.

Roesch, L.F.W., F.A.O. Camargo, F.M. Bento and E.W. Triplett, 2008. Biodiversity of diazotrophs bacteria within the soil, root and stem of field grown maize. Plant Soil, 302: 91-104.

Rolfe, B.G., U. Mathesius, M. Djordjevic, J. Weinman, C. Hocart, G. Weiller and W.D. Bauer, 2003. Proteomic analysis of legume-microbe interactions. Comp. Funct. Genom., 4: 225-228.

Rosswall, T., 1976. The internal nitrogen cycle between microorganisms, vegetation and soil. Ecol. Bull., 22: 157-167.

Rubio, L.M. and P.W. Ludden, 2008. Biosynthesis of the iron molybdenum cofactor of nitrogenase. Ann. Rev. Microbiol., 62: 93-111.

Saikia, S.P. and V. Jain, 2007. Biological nitrogen fixation with non-legumes: An achievable target or a dogma?. Curr. Sci., 92: 317-322.

Schlegel, H.G., 1991. General Microbiology. 7th Edn., Cambridge University Press, Cambridge, pp: 435-445.

Schmidt, W., 1995. Phylogeny reconstruction for protein sequences based on amino acid properties. J. Mol. Evol., 41: 522-530.

Schultz, M. and A. Kondorosi, 1998. Regulation of symbiotic root nodule development. Ann. Rev. Genet., 32: 33-57.

Searls, D.B., 2000. Bioinformatics tools for whole genomes. Annu. Rev. Genomics Hum. Genet., 1: 251-279.

Seefeldt, L.C., I.G. Dance and D.R. Dean, 2004. Substrate interactions with nitrogenase: Fe versus Mo. Biochemistry, 43: 1401-1409.

Selao, T.T., S. Nordlund and A. Noren, 2008. Comparative proteomic studies in *Rhodospirillium rubrum* grown under different nitrogen conditions. J. Proteome Res., 7: 3267-3275.

Sen, S. and J.W. Peters, 2005. The thermal adaptation of the nitrogenase Fe protein from thermophilic *Methanobacter thermoautotrophicus*. Proteins, Struct. Func., Bioinform., 62: 450-460.

Sen, A., S. Sur, A.K. Bothra, D. Benson, P. Normand and L.S. Tisa, 2008. The implication of lifestyle on codon usage patterns and predictably highly expressed genes for three *Frankia* genomes. Anton. Van. Leewen., 93: 335-346.

Sen, A., S. Sur, L.S. Tisa, A.K. Bothra, S. Thakur and U.K. Mondal, 2010. Homology modeling of the *Frankia* nitrogenase iron protein. Symbiosis, 50: 37-44.

Sharp, P.M., T.M.F. Tuohy and K.R. Mosurski, 1986. Codon usage in yeast cluster-analysis clearly differentiates highly and lowly expressed genes. Nucleic Acids Res., 14: 5125-5143.

Sharp, P.M. and W.H. Li, 1987. The Codon Adaptation Index-a measure of directional synonymous codon usage bias and its potential applications. Nucleic Acids Res., 15: 1281-1295.

Sharp, P.M., M. Stenico, J.F. Peden and A.T. Lloyd, 1993. Codon usage: Mutational bias, translational selection, or both?. Biochem. Soc. Trans., 21: 835-841.

Sims, G.E., S.R. Jun, G.A. Wu and S.A. Kim, 2008. Alignment free genome comparison with feature frequency profiles (FFP) and optimal resolutions. Proc. Natl. Acad. Sci., 106: 2677-2682.

Sprent, J.I. and E.K. James, 2007. Legume evolution: Where do nodules and mycorrhizas fit in?. Plant Physiol., 144: 575-581.

Stevenson, F.J., 1972. Nitrogen: Element and Geochemistry. In: The Encyclopedia of Geochemistry and Environmental Sciences, Fairbridge, R.W. (Ed.). Van Nostrand Reinhold publisher, New York, London, pp: 795-801.

Stoffels, M., T. Castellanos and A. Hartmann, 2001. Design and application of new 16S-rRNA-targeted oligonucleotide probes for the *Azospirillum-Skermanella-Rhodocista*-cluster. Syst. Applied Microbiol., 24: 83-97.

Supek, F. and K. Vlahovicek, 2004. INCA: Synonymous codon usage analysis and clustering by means of self-organizing map. Bioinformatics, 20: 2329-2330.

Sur, S., A. Pal, A.K. Bothra and A. Sen, 2005. Moderate codon bias attributed to translational selection in nitrogen fixing genes of *Bradyrhizobium japonicum* USDA110. Bioinformatics Ind., 3: 59-64.

Sur, S., A. Sen and A.K. Bothra, 2006. Codon usage profiling and analysis of intergenic association of *Frankia* EuIK1 nif genes. Ind. J. Microbiol., 46: 363-369.

Sur, S., A. Sen and A.K. Bothra, 2007. Codon usage analysis of some archaeal methanogenic nitrogen fixing genes: relationship to gene expression and biasness. Bioinform. Trends, 2: 47-60.

Sur, S., A.K. Bothra, M. Bajwa, L.S. Tisa and A. Sen, 2008a. *In silico* analysis of *Chlorobium* genomes divulge insights into the lifestyle of the bacteria. Res. J. Microbiol., 3: 600-613.

Sur, S., M. Bhattacharya, A.K. Bothra, L.S. Tisa and A. Sen, 2008b. Bioinformatic analysis of codon usage patterns in a free-living diazotroph, *Azotobacter vinelandii*. Biotechnology, 7: 242-249.

Sur, S., G. Sen, S. Thakur, A.K. Bothra and A. Sen, 2009. *In silico* analysis of evolution in swine flu viral genomes through re-assortment by promulgation and mutation. Biotechnology, 8: 434-441.

Thioulouse, J., S. Doledec, D. Chessel and J.M. Oliver, 1995. ADE Software: Multivariate Analysis and Graphical Display of Environmental Data. In: Sofware per lAmbiente, Guariso, G. and A. Rizzoli (Eds.). Patron, Bolonia, pp: 57-61.

Udvardi, M.K., 2002. Legume genomes and discoveries in symbiosis research. Genome. Biol., 3: 1-3.

Umashankar, V., V. Arun Kumar and D. Sudarsanam, 2007. ACUA: A software tool for automated codon usage analysis. Bioinformation, 2: 62-63.

Villalobos, A., J.E. Ness, C. Gustafsson, J. Minshull and S. Govindarajan, 2006. Gene designer: A synthetic biology tool for constructing artificial DNA segments. BMC Bioinformatics, 7: 285-285.

Wade, N., 2000. Genetic Code of Human Life is Cracked. Times, New York.

White, J., J. Prell, E.K. James and P. Poole, 2007. Nutrient sharing between symbionts. Plant Physiol., 144: 604-614.

Willey, R.W., 1979. Intercropping: Its importance and research needs. Part 2. Agronomy and research approaches. Field Crop Abstr., 32: 73-85.

Williams, R.J.P., 2007. A systems view of the evolution of life. J. R. Soc. Interface, 4: 1049-1070.

Winogradsky, S.N., 1893. Sur lassimilation de lazote gazeux de latmosphere par les microbes. C. R. Acad. Sci., 116: 1385-1388.

Wright, F., 1990. The effective number of codons used in a gene. Gene, 87: 23-29.

Wu, G., D.E. Culley and W. Zhang, 2005. Predicted highly expressed genes in the genomes of *Streptomyces coelicolor* and *Streptomyces avermitilis* and the implications for their metabolism. Microbiology, 151: 2175-2187.

Wu, G., N. Bashir-Bello and S.J. Freeland, 2006. The synthetic gene designer: A flexible web platform to explore sequence manipulation for heterologous expression. Protein Expr. Purif., 47: 441-445.

Yang, G.P., F. Debellé, A. Savagnac, M. Ferro and O. Schiltz *et al.*, 1999. Structure of the *Mesorhizobium huakuii* and *Rhizobium galegae* Nod factors: A cluster of phylogenetically related legumes are nodulated by rhizobia producing Nod factors with alpha,beta-unsaturated N-acyl substitutions. Mol. Microbiol., 34: 227-237.

Young, J.P., L.C. Crossman, A.W. Johnston, N.R. Thomson and Z.F. Ghazoui *et al.*, 2006. The genome of *Rhizobium leguminosarum* has recognizable core and accessory components. Genome. Biol., 7: R34-R34.

Young, J.P.W., 1992. Biological Nitrogen Fixation. Chapman and Hall, New York, pp: 43-86.

Zehr, J.P. and L.A. McReynolds, 1989. Use of degenerate oligonucleotides for amplification of the nif H gene from the marine cyanobacterium *Trichodesmium* spp. Applied Environ. Microbiol., 55: 2522-2526.

Zehr, J.P., D. Harris, B. Dominic and J. Salerno, 1997a. Structural analysis of nitrogenase iron protein: Implications for aerobic nitrogen fixation ability. FEMS Microbiol. Lett., 153: 303-309.

Zehr, J., M.T. Mellon and W.D. Hiorns, 1997b. Phylogeny of cyanobacterial nif H genes: Evolutionary implications and potential applications to natural assemblages. Microbiology, 143: 1443-1450.

Zehr, J.P., M.T. Mellon and S. Zani, 1998. New nitrogen-fixing microorganisms detected in oligotrophic oceans by amplification of nitrogenase (nifH) genes. Applied Environ. Microbiol., 64: 5067-5067.

Zehr, J.P., E.J. Carpenter and T.A. Villareal, 2000. New perspectives on nitrogen-fixing microorganisms in tropical and sub-tropical oceans. Trends Microbiol., 8: 68-73.

Zehr, J.P., B.D. Jenkins, S.M. Short and G.F. Steward, 2003. Nitrogenase gene diversity and microbial community structure: A cross-system comparison. Environ. Microbiol., 5: 539-554.

Zheng, L., D.L. Cash, D. H. Flint and D. R. Dean, 1998. Assembly of iron-sulfur clusters. Identification of an iscSUA-hscBA fdx gene cluster from *Azotobacter vinelandii*. J. Biol. Chem., 273: 13264-13272.