Conclusion

The major findings obtained from the study are as follows:

1. Synonymous codon usage analysis in the studied diazotroph genomes divulged widespread variation at the intraspecific and intraspecific levels. The results showed considerable degree of heterogeneity. Interplay of different factors like mutational pressure, translational efficiency, gene expression levels, GC3 compositional bias etc. manipulated codon usage variation and provided insight into the nature of diazotroph genomes. Correspondence analysis in majority of the organisms showed that most of the genes are confined to a core in the plot. Using the codon adaptation index, potentially highly expressed (PHX) genes were predicted and their features scrutinized. The results provided estimation of the global gene expression patterns in the studied diazotrophs enhancing the knowledge on the characteristics of diazotrophs. Functional analysis of the PHX genes highlighted their influence on the lifestyle patterns of the nitrogen fixing microorganisms. Majority of the PHX genes are associated with metabolic functions
2. Variation of codon usage patterns for nitrogen fixation related genes in the different bacterial groups were due to a combination of a number of factors. They were heterogeneous in nature. Correspondence analysis revealed the conserved nature of the genes. The nitrogen fixation related genes had high CAI values implying higher expression levels. Identification of a number of core nitrogen fixation related genes in the PHX category signified their role in influencing the lifestyle of the organisms in their respective environmental conditions.
3. Correlation of codon usage bias with tRNA content portrayed that fast growers had high number of tRNA and had a small number of anticodons

implying that they possessed less diverse tRNAs. Organisms having reduced genome had less number of tRNA genes. In the studied organisms where ribosomal protein genes showed higher codon bias varying widely from the protein coding genes of the organism, translational efficiency plays a vital role in effecting codon usage variation.

4. TTA codon containing genes in the studied GC rich organisms revealed diversity. TTA codon containing genes are under the influence of mutational pressure and are less biased in their codon usage patterns. They had a number of genes in the PHX category. Functional analysis revealed that COGs linked to metabolism had the highest number of PHX genes implying the role played by TTA containing genes in influencing the metabolism in the respective habitats and ecological niches of the organisms.

5. The proteomes of nitrogen fixing organisms show diversity at the interspecific and intraspecific level. The proteomes showed a tri-modal distribution for pI. Nitrogen fixation related proteins are more acidic in contrast to whole proteomes. Analysis of pI across COGs throws light upon the variation and lifestyle of organisms and substantiates the role of isoelectric point as a molecular signature. Proteomes of the studied organisms show a shift towards acidity and are influenced by habitat and environmental adaptations. Higher AAAI values of proteomes imply their better adaptability in shaping frequency of amino acids. Hydropathicity and aromaticity is relevant for evaluating amino acid usage in these organisms and variations across proteomes.

6. The findings for condensed matrix based phylogeny for *nif*H, D & K genes, and whole genomes from cyanobacteria, proteobacteria, clostridial members, actinobacteria, green sulfur bacteria and archaea indicated that structural properties of DNA sequences are guided by different descriptors and invariants. Results support 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 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 & K phylogenies, with respect to horizontal gene transfer and gene duplication events. High

degree of heterogeneity present among *nif* genes implied action of mutation and selection pressures with unlike intensities. The 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.

7. *nif* genes did not evolve as an unit.

8. The NifH protein structures presented for *Frankia* strains are reliable offering insights into the 3D structural framework as well as structure-function relation of NifH protein. 4Fe-4S cluster represented the core functional region. Thiol ligands controlled conformational reaction of the protein. Active site pockets on the protein surface and amino acid arrangement inside it, crafts the inherent physico-chemical features vital for NifH protein functioning. Clefts and cavities contained biologically important residues. *In silico* site-directed mutagenesis results revealed that mutations in functional residues hamper nitrogen fixation. Accessible surface is solvent accessible. Negatively charged and polar uncharged residues are present on the outermost surface. RMSD analysis substantiated sequence similarity result. Normal mode analysis revealed the rigidity of the model. Validation results proved that stereochemical properties and structure quality are reliable with their features.