Abstract

Actinobacteria are mostly Gram +ve high G+C containing microbes representing one of the largest bacterial phyla. They are ubiquitously distributed between aquatic and terrestrial niche and generally have mycelial lifestyle with complex morphological differentiation. They constitute a major source of naturally derived antibiotics with anticancer, anthelmintic, and antifungal compounds which have vital implications in clinical use. Consequently, these bacteria have significant role in the field of biotechnology, agriculture and medicine. Moreover, the host (both plant and animal) associated actinobacteria are well adapted to their lifestyle and make diverse association (both pathogenic and advantageous) with their respective hosts. Hence, according to the diverse niche adapted by different Actiobacterial genera, they can be classified into several groups. For instance, animal host associated, plant host associated, soil inhabitants, aquatic, thermal and extremophiles (other than thermal). Development of next generation sequencing techniques has advanced the biological research area providing the high-throughput data. Since the NGS data are always big-data, manual analysis is next to impossible. Moreover, manual data mining and analysis may results into tremendous error misleading the actual fact. Here comes the role of Bioinformatics which with the aid of programming, computational methodology and statistics can deal with these big-data easily and quickly.

In this study, we tried to cover the genomic and proteomic aspects acquired by actinobacteria for their metabolism, niche adaptation and signaling. We tried to cover most of the genera of Actinobacteria whose whole genome sequencing are done and are available in the public domain database. While studying the metabolomics of considered organisms we found there are some critically important metabolic pathways where a higher number of potentially highly expressed proteins (and their encrypting genes) were involved. Those pathways were carbohydrate

metabolism, amino acid metabolism, energy metabolism, lipid metabolism, co-factors and vitamin metabolism, terpenoid and polyketide metabolism, nucleotide metabolism and xenobiotic metabolism. The major indices with pivotal significance on the genomic and proteomic attributes on the genes and related proteins involved in the aforementioned metabolic pathways were examined. Extensive studies on each of the parameters which may govern their codon usage prototype revealed compositional constrain, optimal codons, translational efficiency and natural selection as major indices with substantial effect over the selected organisms. Further, the amino acid usage pattern and protein biosynthetic energy was also considered. This displayed a very interesting pattern of aromatic amino acid usage variation with protein expression level. Comparative along with phylogenomics study proved to be major sources for studying genomic features of investigated organisms. We tried to correlate the results obtained from the comparative genomics study with those of phylo-genomic investigation to further validate and corroborate our outcome.

Evaluation of the characteristic features as well as evolutionary strategies taken up by the secretory proteins, which are mainly involved in signal transduction, was another main aspect of this thesis. A thorough study on the secretomes of some *Mycobacterium* revealed a distinct difference among the signaling proteins of pathogenic and non-pathogenic strains. Consequently a higher evolutionary rate of secretomes and their consequent signalP parts were exposed. This further implicates the substantial role of signalP (signaling proteins) on the niche adaptation. Further, the secretome set was also found to be positively correlated to the proteome size of the considered *Mycobacterium* strains indicating towards the fact that magnitude of signaling enhance with complexity of life. The secretomes of extremophilic stone dwellers (*Geodermatophilus*, *Blastococcus* and *Modestobacter*) were also studied. This study revealed their importance in harsh condition. Further, the secretomes were found to be evolved faster than the non-secretomes. Since, the considered extremophilic strains are stone dwellers and inhabit in harsh condition the faster evolution of secretomes may favour the modification of signaling with their circumstances ultimately giving a

beneficiary effect on them.

Another aspect of this thesis work was to obtain a knowledge regarding significance of Carbohydrate- activated enzymes (CAZymes) in host-microbe interaction of *Frankia* and its associated hosts. CAZymes are basically important for braking down the complex carbohydrates or polysaccharides. *Frankia* through their CAZymes can easily break down the plant polysaccharides (obtained from either host plant or soil) into simple monosaccharides and utilize them. The CAZymes were also found to be potentially highly expressed suggesting their importance among *Frankia*. The evolutionary analysis revealed a lower evolutionary rate of CAZymes than rest of the genes implying the persistence of natural selection over them.

Along with these aforementioned aspects of Actinobacteria their protein domains were also considered. Protein domains are semiautonomous parts of proteins having their own independent structure and function. Domain-domain interaction is also very important in maintaining a proper and balanced lifestyle of each organism. Moreover, presence and absence of a particular domain may depend on the niche of the bacteria. We tried to build a phylogenetic tree with information based on the presence and absence of domains among selected Actinobacteria. The domain based phylogeny was compared with a conventional phylogeny. Some genera were recovered perfectly from the domain based phylogeny whereas; a few were behaving in a different way. It was found that, those organisms (with unusual clustering pattern) grouped with taxonomically different genera belonging to same niche. This result indicated towards a probability that, habitat may control the presence and absence of protein domains in microbes.

Another important aspect of protein domains is their own unique tertiary structure. Through literature studies we found that Per-Arnt-Sim (PAS) domain is very crucial in maintaining the overall signal transduction machinery in more or less all living beings. However, the role of PAS in Actinobacteria was not thoroughly studied. Hence we took up an approach to reveal the importance of tertiary structure of PAS domain in modulating the biological network. We found a very interesting result predicting that, interaction of a domain with its co-

domain play a significant role in maintaining the signaling network system. Thus, the domain-domain interaction may act as a crucial factor in regulating the biological network or vice versa.

The last aspect of this thesis was a bit different from the rest. Since, *Mycobacterium tuberculosis* is some of the deadliest pathogens causing enumerate deaths worldwide along with India and several medicinal trials are failing to eradicate them, we identified some novel drug targets from one Indian *M. tuberculosis* strain whose whole genome sequence is already available in public domain database. These predicted proteins were further validated with molecular docking study. We selected already known anti-TB agents as ligands and docked them with aforementioned proteins. Molecular study revealed that, those predicted targets can be used as potential target proteins in the battle against TB. However, this can only be confirmed totally after a series of experiments.