Chapter 2
Review of Literature

Facts are many, but the truth is one.

-Rabindranath Tagore

2.1 Features of Actinobacteria

The phylum Actinobacteria represents one of the largest taxonomic units among major lineages currently recognized within Bacteria (Stackebrandt et al. 2000, Stackebrandt et al. 1997) in terms of number and variety of identified species. It comprises gram-positive high G+C content bacteria. The G+C content generally ranges from 51% (some *Corynebacterium*) to more than 70% in *Streptomyces* and *Frankia*. An exception is the genome obligate pathogen *Tropheryma whipplei*, whose G+C content is less than 50%. Actinobacteria exhibit a broad range of morphologies, from rod-coccoid (*Arthrobacter*) or coccoid (*Micrococcus*) to fragmented hyphal forms (*Nocardia* spp.) or highly differentiated branched mycelium (*Streptomyces* spp.) (Atlas et al. 1997).They also exhibit diverse metabolic and physiological properties, such as the formation of a wide variety of secondary metabolites and production of extracellular enzymes (Schrempf et al. 2001). Some of these secondary metabolites especially produced by *Streptomyces* species have been proved to be potent antibiotics (Lechevalier et al. 1967). This trait turned *Streptomyces* into primary antibiotic-producing organisms exploited by the pharmaceutical industry (Bérdy et al. 2005).Moreover, diverse forms of lifestyles are encountered among Actinobacteria.For instance pathogens (*Mycobacterium*, *Nocardia*, *Tropheryma*, *Corynebacterium*, *Propionibacterium* etc.), plant commensals (*Leifsonia*), nitrogen-fixing symbionts (*Frankia*), soil inhabitants (*Streptomyces*), thermal (*Acidothermus*), stone dwellers (*Geodermatophilus*) and gastrointestinal tract (GIT) inhabitants (*Bifidobacterium* spp.).

Abnormal developmental features are also displayed by many actinobacterial genera, for example,the continual non-replicating state exhibited by certain mycobacteria or sporulating aerial mycelium formation in *Streptomyces* species. Actinobacteria are extensively distributed equally in terrestrial and aquatic (including marine) ecosystems, specially in soil, where they play a vital role in the recycling of unruly biomaterials by decomposition as well as humus formation (Goodfellow et al. 1983, Stach et al. 2005). Furthermore, some *Bifidobacterium* species are used as active component in a variety of so-called functional foods due to their apparent health-promoting and probiotic properties. They have been reported to take part inbile salt hydrolase activity, protection against pathogens mediated through the process of competitive exclusion, immune modulation along with the ability to stick to mucus or intestinal epithelium (Lievin et al. 2000, Ouwehand et al. 2002, Stanton et al. 2005). Thus actinobacteria may be relevant to human and veterinary medicine, biotechnology, and ecology. Moreover the genomic heterogeneity observed from already sequenced actibacterial genomes reflects their biodiversity.

2.2 Evolution and Dynamics of Bacterial Genomes

The primary genetic proceedings that establish genome shape and structure are believed to be (1)horizontal gene transfer (HGT), (2) gene duplication, (3) chromosomal rearrangements and (4) gene loss. Despite several efforts to quantify the relative contribution of each of these processes, no consistent model can so far explain and outline the evolutionary development of bacteria based on their existing genome structure (Andersson et al. 2000, Hong et al. 2004, Kunin et al. 2003, Snel et al. 2002). A brief review regarding all the aforementioned processes could be helpful in reconsidering the genomic aspects of organisms.

2.2.1 HGT:

The introduction of alien or novel genes by HGT permits for quick niche-specific adaptation, which may lead to bacterial diversification and speciation (Cohan et al. 2001). Bacterial genome evolution is dependent upon the combined outcome of acquired genes through cell division (vertically inherited) and by HGT (Woese et al. 2000). Taking this perception to its extreme, one can argue that two

bacterial taxa are more linked to each other than to a third one not because they allocateto a more recent ancestor but because they swap over genes more frequently (Gogarten et al. 2002). HGT is responsible for increasing the competitiveness among bacteria in their natural habitats. For instance, in some pathogenic bacteria, DNA segments with many virulence genes and gene clusters (pathogenicity islands), appear to be acquired by HGT (Ochman et al. 2000). Actinobacterial examples of virulence genes transmission through HGT are rare (Rosas-Magallanes et al. 2006). Of these, three cases showed to signify obvious HGT events: (i) presence of major diphtheria toxin gene in phages of *Corynebacterium diphtheria*, (ii) *Mycobacterium ulcerans* carries the macrolide toxin genes liable for the ulceration in their linear plasmid (Stinear et al. 2004), (iii) a large chromosomal segment of *Streptomyces turgidiscabies* associated with causing potato scab can be transferred via conjugation (Loria et al. 2006). In addition, it has been argued that the virulence operonof *Mycobacterium tuberculosis* Rv0986-8 which plays a vital role in host phagocytic cells parasitism by enhancing the ecological fitness of the infecting mycobacterium was acquired by HGT (Pethe et al. 2004). Other genetic studies of the ancestral *M. prototuberculosis* species have specified that various HGT events happened before the evolutionary bottleneck that led to the appearance of the *M. tuberculosis* complex (Gutierrez et al. 2005), perhaps from the Indian subcontinent (Filliol et al. 2006).

Bioinformatic methods to recognize HGT events are based predominantly on the analysis of deviation in the GC content (GC deviation), four-letter genomic signatures, dinucleotide differences, and/or codon usage. Some geneticists often proposed HGT as the explanation for genes present in only one organism which would mean that HGT frequency is pretty low (below 10% of the total gene complement) (Kunin et al. 2003, Snel et al. 2002). Interestingly, a recent analysis revealed that many actinobacteria specific proteins were also encoded by *Magnetospirillum magnetotacticum* which is an alpha-proteobacterium, but not by any other alpha-proteobacterial strains. This proposes that *M. magnetotacticum* may have acquired those genes from actinobacterial species by HGT (Gao et al. 2006).

An interesting case of HGT between

Chlamydia and some Actinobacterial genera (e.g., *Streptomyces, Tropheryma, Bifidobacterium, Leifsonia, Arthrobacter,* and *Brevibacterium*) have already been described (Griffiths et al. 2006). In serine hydroxymethyltransferase enzyme (GlyA protein), two conserved inserts of 3 and 31 amino acids remain present in some chlamydiae along with the aforementioned subset of Actinobacteria. Likewise, these bacteria also contain a preserved 16-amino-acid insert in the enzyme UDP-N-acetylglucosamine enolpyruvyl transferases (MurA) responsible for peptidoglycan biosynthesis. However, the physiological and functional significance of these HGT events between Actinobacteria and Chlamydiae is presently unclear.

2.2.2 Gene decay:

Bacterial genome size is dogged by the outcome of numerous opposing forces. Genome contraction occurs by deletion bias and genetic drift whereas, a prominent selection pressure on the functional genes encourages the preservation of DNA.Genome augmentation depends on both acquisitions of alien DNA and gene duplications, coupled to adaptive benefits (Mira et al. 2001). DNA loss may be either large deletions that span multiple loci or deletions of one or a few nucleotides (Andersson et al. 2001). The influence of these altered routes is erratic among bacterial lineages (Mira et al. 2001). Inactivation and deleterious mutations in genes with little involvement to fitness can be transmitted to progeny and mount up in populations, ultimately leading to gene loss; whereas such mutations in genes that are lethal will prevent the production of progeny and so will be eradicated from populations, resulting in the conservation of the functional gene (Ochman et al. 2006).

Gene inactivation and loss are mainly apparent in several bacterial groups with host-associated lifestyle, in which the host provides many of the metabolic intermediates, thereby preventing the requirement to preserve many biosynthetic genes. In endosymbiotic bacteria, such as *Buchnera* and *Rickettsia*, loss of individual loci or operons is the only resource of discrepancy in the gene inventories between species (McLeod et al. 2004, Tamas et al. 2002). A clear example of genome reduction is provided by *Mycobacterium leprae*, which has lost more than 1,000 genes compared with *M. tuberculosis* (Cole et

al. 2001). Moreover, the presence of non-functional genes, i.e., pseudogenes, in *M. leprae* suggests that this genome loss is still in progress. Although the principles for pseudogenes identification differ among studies, the overall rationale is quite similar: the predicted protein must be distorted to degree that abolishes its function. The thresholds applied for identifying pseudogene are based on the organization and known size of functional domains within proteins, the observed length deviation within individual gene families, and available data on experimentally disrupted proteins (Ochman et al. 2006). Generally, pseudogenes include cases where deletion of functional part or insertion of stop codons have resulted in distorted protein with very small gene length (less than 80% of original length) or where a frameshift or insertion has altered more than 20% of the amino acid sequence of the protein (Lerat et al. 2004). Majority of the annotated bacterial pseudogenes are among the open reading frames (ORFs) with unknown functionality. The lack of common pseudogenes among different strains of same species indicates that pseudogenes are produced continually, are eliminated quickly, and thus only rarely persevered in bacterial genomes (Ochman et al. 2006). Other bacteria reveal a lower level of gene loss: in *Rickettsia prowazekii,* anobligate intracellular pathogen where only 12 pseudogenes were identified (Andersson et al. 1998); and a current genome analysis of two *Streptococcus thermophilus* strains (Bolotin et al. 2004) found that pseudogenes constituted only 10% of the total genome, reflecting adaptation of *S. thermophilus* to their specific environment which is milk (Bolotin et al. 2004).

2.2.3 Gene duplications:

Earlier it was thought that bacterial genomes evolved from a much smaller ancestral genome through a number of gene duplication events and the resulting generation of paralogs (Kunisawa et al. 1995). However, some analysis based on currently available data does not support this theory and reveals that gene duplications contribute only diffidently to genome evolution (Coenye et al. 2005). Despite this, it has been noted that genes participated in a precise adaptation have been conserved after duplications signifying that gene duplication does have an evolutionary role (Coenye et

al. 2005). This is nicely exemplified by the mycobacterial paranome, corresponding to a functional class of genes concerned with fatty acid metabolism, in accordance to the composite nature of the mycobacterial cell wall and possibly reflecting the adaptive evolution of microbial cellular structure (Coenye et al. 2005, Tekaia et al. 1999).

2.2.4 Genome rearrangements:

Along with the afore-mentioned factors, chromosome rearrangements can also participate in genome arrangements. Synteny, a term used for indicating the preservation of gene order between genomes, can be used as a phylogenetic tool to examine relationships among species, since the degree of genome rearrangements enhances linearly in relation to the bacterial divergent time (Korbel et al. 2002, Wolf et al. 2001). Chromosomal rearrangement is chiefly dependent on the repeated and mobile elements activity for example functionality of transposons, insertion sequences (ISs), prophage sequences, and plasmids (Kolsto et al. 1997). Bacterial genomes with a higher density repeat sequence have higher rearrangement rates leading to a rapid loss of gene order (Rocha et al. 2003). Homologous recombination between these repeated sequences catalyze genetic rearrangement as well as gene loss leading to taxonomic diversification. These recombination events probably have endorsed in *T. whipplei* speciation (Raoult et al. 2003). Moreover, evolution of chromosome is influenced by bulky chromosomal rearrangements, e.g., large inversions (Eisen et al. 2000).

2.3 Taxonomic reclassification of Actinobacteria

Ray fungi, also called as Strahlenpilze by Lieske (1921), or actinobacteria with a tricky taxonomy (Krassilnikov 1941) were historically considered as ‘intermediary’ between bacteria and fungi. The name itself comes from the colony morphology on agar exhibiting radial growth, a characteristic common to fungi. The mycelial shape and musty smell of actinobacteria is also similar to fungi. Moreover, the peptidoglycan based cell wall of this group is also a property shared to Firmicutes. The very primitive taxonomical treatment of actinobacteria was proposed by Buchanan (1917) who proposed an order name Actinomycetales. Some taxonomical works have been based on the perception of evolutionary trend from a simple form to more

complicated features like hyphae and sporangia (Kluyver & van Niel 1936). However most of the taxonomists were not satisfied since morphological features were not considered which may result in faulty positioning, especially in classifications specifically dependent upon a dichotomous scheme.

Chemical criteria were initially very few and could not provide a solid base for taxonomy, yet they have gathered a solid set of data compensating for few lost functions. Now-a-days, to explain a species, it is compulsory to afford a measure G+C content of DNA which is a characteristic of the genome; phospholipids analysis which is an important feature of membrane; the di-amino acids characterising cell wall; or quinones which represent the respiratory chain of organisms. None of these elements taken independently is adequate to recognize a microbe; however, taken as a whole they can give way a solid taxonomic basis in amalgamation with growth characteristics and morphology. Sokal & Sneath (1963) used a mixture of such characters as the basis of numerical taxonomical study. These approaches have allowed the classification of several actinobacterial lineages (Goodfellow 1967), especially the genera *Streptomyces* (Williams et al. 1983) and *Mycobacterium* (Tsukamura et al. 1969).

The arrival of sequencing of 16S rRNA, 16S rRNA gene classification (Stackebrandt & Woese 1981) and advanced molecular phylogenetic tools have advnced the search for a molecular clock that is required to be present in all lineages. These techniques presented actinobacteria as a consistent subdivision of 'Gram-Positive bacteria' along with the firmicutes (low G+ C), *Bacillus* and *Clostridium* (Woese 1987). This confirmation supported Actinobacteria as coherent, with little genetic distances.

Actinobacteria are generally aerobic and posses filaments (in branching pattern) except Bifidobacteriales (Woese 1987). This treatment was reserved in many consecutive editions of Bergey's Manual and the class name was available in 1997 (Stackebrandt et al. 1997) although untenably because no type order was proposed at the time (Euze´by & Tindall 2001). However, 16S rRNA genes sequences and phylogenetic reconstructions may not be considered as the golden benchmark in bacterial taxonomy for several

reasons. First, there are examples of species with more than one copy of the 16S rRNA gene differing by approximately 6%, as in *Thermomonospora chromogena* (Yap et al. 1999), which was taken as support of lateral transfer. Secondly, the 16S rRNA gene may be plasmid-borne as in *Bacillus megaterium* (Kunnimalaiyaan et al. 2001), further following the idea of lateral transfer. Finally, the 16S rRNA gene is a single marker that thus disobeys one basic principle of biology, that any investigation should be considered for reproducibility. This constraint led to multi-locus sequence analysis (MLSA) developed originally for *Neisseria* (Maiden et al. 1998), where five or more preserved genes were sequenced and analysed. This approach is now commonly used, mainly to distinguish species within a genus. It helped the re-characterization of the genera *Nocardia* (McTaggart et al. 2010) and *Streptomyces* (Doroghazi & Buckley 2010), with some noteworthy deviations from the topological pattern obtained by 16S rRNA gene sequences, indicating extensive recombination. The most current taxonomical treatment of Actinobacteria (Ludwig et al. 2012a, b), based fundamentally on the 16S rRNA gene phylogeny, has noticeably tailored all levels of their taxonomy. There are six classes comprising of five basal ones (*Acidimicrobiia, Coriobacteriia, Nitriliruptoria, Rubrobacteria* and *Thermoleophilia*) each with one or two orders. The main class *Actinobacteria* comprises of 15 orders (*Actinomycetales*, *Actinopolysporales, Bifidobacteriales*, *Catenulisporales*, *Corynebacteriales*, *Frankiales, Glycomycetales*, *Jiangellales' Kineosporales*, *Micrococcales*, *Micromonosporales*, *Propionibacteriales*, *Pseudonocardiales*, *Streptomycetales Streptosporangiales*).

Gao and Gupta (2012) proposed an approach of concatenating 35 proteins to re-classify actinobacterial phylogeny, with an idea of indels to corroborate the major lines of that classification (Wu et al. 2009). Study of Verma et al. (2013) revealed the need of reassessment in the orders *Frankiales* and *Micrococcales.* Sen et al. (2014) performed a detailed phylogenetic reconstruction exclusively for Actinobacteria using both bootstrapping (Felsenstein 1985) MLSA and Prunier approach (Abby et al. 2010) and spitted *Frankiales* into

Frankiales ord. nov., *Geodermatophilales* ord. nov., *Acidothermales* ord. nov. and *Nakamurellales* ord. nov.. Moreover, *Micrococcales* was also divided into several monophyletic orders – *Micrococcales (Micrococcus, Kocuria, Rothia, Arthrobacter, Tropheryma, Microbacterium, Clavibacter* and *Leifsonia); Cellulomonales (Beutenbergia, Cellulomonas, Xylanimonas, Jonesia* and *Sanguibacter); Brachybacteriales (Brachybacterium) and Dermacoccales (Kytococcus, Intrasporangium).*

2.4 Interaction with niche

Actinobacteria being one of the major representatives of rhizospheric microbial flora play diverse roles in plant-associated microbial communities. Some genera are predominant soil saprophytes while others are either symbionts or endophytes or plant pathogens.

2.4.1 Actinobacteria as nitrogen fixers

Frankia, a genus with the ability to fix nitrogen, can live both as soil saprophytes as well as endophytic symbionts and are competent to establish mutualistic symbiotic association with non-leguminous plants notably *Alnus*, *Casuarina* and *Elaeaginus* allowing their growth in nitrogen-poor condition (Benson and Silvester 1993; Chaia et al. 2010; Schwencke and Carú 2001). Elucidation of the *Frankia* genomes has revealed novel possibilities in metabolic diversity, stress tolerance and natural product biosynthesis, which aid in establishment of *Frankia*-actinorhizal symbiosis (Schwencke and Caru 2001; Benson and Dawson 2007). Depending upon the host specificity, *Frankia* has been clustered in four major groups (Ghodbhane-Gtari et al. 2013, Nouioui et al. 2011; Normand et al. 1996; Cournoyer and Lavire 1999, Ghodhbane-Gtari et al. 2010). Members of lineage I infect plants of the Betulaceae, Myricaceae, and Casuarinaceae families (except Gymnostoma). Dryadoideae (all actinorhizal Rosaceae), Coriariaceae, Datiscaceae, and the genus Ceanothus (Rhamnaceae) are infected by strains of lineage II. The lineage III stains are most promiscuous infecting Elaeagnaceae, Colletieae (all actinorhizal Rhamnaceae except Ceanothus), Myricaceae, Gynmmostomam (Casuarinaceae), and occasionally *Alnus*. On the contrary, Frankia of lineage IV are 'atypical' strains either unable to infect plant or

produce ineffective root nodules which do not participate in nitrogen fixation (Tisa et al. 2016).

Several genes have been reported to be involved in nitrogen fixation. Among them, nitrogen fixing (nif) genes are most important (Normand et al. 2007). Strains of cluster I, II and III own their nif genes however, in lineage IV this gene is totally absent which probably lead to their inability to fix nitrogen. Eleven- twelve nif genes remain clustered together as an operon on 11.6 -14 kb region and generally one copy of each nif gene is present in that cluster (Tisa et al. 2016). However, exceptions are there. For instance, nifV gene of lineage III strains has been reported to be present at a distant from the nif cluster and two copies of nifU were reported in lineage II. Interestingly, six non-nif genes (orf1, orf2, hesA orf3, orf4 and fdx) were also found to be located between nif genes (Tisa et al. 2016). Another important part of frankial N_2 fixation is the presence of specialized thick-walled vesicles (Noridge and Benson 1986; Tisa and Ensign 1987; Normand et al. 2014). Their formation is controlled by partial pressure of oxygen. Vesicles play a pivotal role in protecting nitrogenase from oxygen inactivation (Murry et al. 1985). A natural pentacyclic triterpenoid lipid called hopanoid formed a multilayed envelop surrponding the vesicle structure which act as the oxygen barrier (Berry et al. 1991, 1993). In *Frankia* hopanoid remain as bacterioho -panetetrol and bacterio-hopanotetrol phenylacetate monoester derivatives. Generally, there are two pathways for hopanoid biosynthesis viz. mevalonate pathway and methyl erythritol phosphate (MEP) pathway. However, in *Frankia*, only MEP pathway has been identified (Rohmer 2008). Surprisingly, hopanoid biosynthetic gene cluster have been identified in all *Frankia* strains including non nitrogen fixer lineage IV which indicates that their functional implication is nitrogen independent (Dobritsa et al. 2001; Udwary et al. 2011; Richau et al., 2013, Tisa et al. 2016). Hence, these genes are not only crucial for nitrogen fixation besides, they are also important as primary cell wall component conferring the membrane integrity.

Concurrently, Ni-Fe hydrogenase or uptake hydrogenase (hup) operons are also imperative for Frankia life style. There are two hup operons in lineage I and II strains Among them, operon I is

expressed under free living condition whereas, operon II is expressed under symbiotic condition (Leul et al. 2007; Richau et al. 2013). However, all hup genes were found to be clustered in one operon for lineage III and IV strains. Truncated hemoglobin genes, which are broadly classified into three groups - group I (HbN), group II (HbO), and group III (HbP) act either as NO scavengers or protect nitrogenase from oxygen inactivation by buffering the oxygen level (Wittenberg et al. 2002). Blastp analysis revealed the complete absence of canonical nodABC gene (NodA-acyl transferase, NodB-chitin deacetylase, NodC-chitin synthase) cluster among *Frankia* strains (Tisa et al. 2016) except *Candidatus* Frankia Datisca Dg1(Persson et al. 2015). There prevail two hypotheses behind this finding. First one is, Dg1 follow an identical symbiotic pathway to legume-rhizobium relation and second one is, while sequencing, Dg1 has incorporated genes from other organisms as a part of contamination. Actinorhizal nodules are well known to contain several bacteria within them hence, genes of other microbes can be accidentally mixed to the genome of Dg1 resulting the exceptional presence of nod genes within the genome (Ghodhbane-Gtari and Tisa 2014). However, Chabaud et al. 2015 has revealed a novel signalling mechanism for symbiosis in *Frankia*. Host plant induced *Frankia* have been reported to exert a signal triggering the expression of reinfection genes in host plant as well as $Ca^{2}+$ level. Their additive effect results in biologically active symbiotic signal. Moreover, next-generation proteomic analysis has also provided plethora of information regarding the proteome footprinting of frankial metabolomics altered by root exudates of compatible and non-compatible host plants. This study revealed that, proteins having functionality in the early steps of host recognition and infection were expressed highly in *Frankia* induced by root exudate of compatible host plant (Ktari et al. 2017).

Bioinformatic analysis has also explored a relation between the genome size and niche specificity as well as biogeographic ranges among this genus. For instance, the genome size of narrow-host-range- *Casuarina* strain *Frankia* sp. CcI3 is 5.4 Mb whereas, the size expands to 8.9 Mb for *Frankia* sp. Ean1pec which is a broad-host-range *Elaeagnus* strain as well as soil cosmopolitan. The

infectivity of *F. alni* ACN14a is restricted mainly to *Alnus* and Myricaceae hence, they possess a genome with intermediate size (Tisa et al. 2016). Thus a hypothesis prevailed that, the genome size of Frankia changes with their host range and it held true with all the strains of this genus sequenced so far. Furthermore, study with horizontal gene transfer (HGT) and insertion elements (IS) in *Frankia* has availed us with information regarding their genomic plasticity (Tisa et al. 2016). *Frankia* sp. Eul1c appeared to be the most stable genome with a very few HGT and IS elements. Thus, our understanding regarding the infection process of Frankia could be enhanced through a detailed comparative study.

2.4.2 Actinobacteria as plant protector and plant defense elicitor

Actinomycetes are well recognized for their capacity to produce a broad range of metabolites which can act against plant pathogens (Van et al. 2008). The mechanisms by which actinobacteria inhibit the pathogens include antibiosis, nutrient competition, quorum quenching, nitrous oxide and degradative enzyme production (Conn et al. 2008). Some strains also produce siderophores with the ability to chelate iron and thus deprive other microbes from this essential nutrient (Jog et al. 2014). Some water soluble chitinolytic enzymes produced from soil actinobacteria are also effective against fungal infection and soil borne pathogens. Broad range of antibiotics produced by this particular microbial population is well recognized plant protectors against pathogenic bacteria, oomocytes, nematodes and fungi (Van et al. 2008). Moreover, production of nitrous oxide by some *Streptomyces* strains elicits the plant defense mechanism. Streptomycetes can also degrade a signaling molecule responsible for pathogenicity in *Pectobacterium carotovorum* (Baz et al. 2012). Thus, actinobacteria can act as a soldier of plant defense mechanism.

2.4.3 Actinobacteria help in plant growth:

Several nitrogen fixing actinobacteria have been found to be associated with the plant root system. Majority of them aid in the development of plant (Qin et al. 2011). For example, *Corynebacterium* sp. AN1, which was isolated from forest phyllosphere can reduce acetylene thus substituting the role of nitrogenous fertilisers and helping in plant growth promotion

(Nimaichand, 2016). *Pseudonocardia dioxanivorans* CB1190 has also been reported to fix dinitrogen (Mahendra et al. 2005). Above all, *Streptomyces*, accounting an abundant percentage of soil microflora and an effective plant colonizer is well recognized as source of plant growth promoters and other biocontrol agents (Igarashi et al. 2002). Besides, endophytic (microbes remain within the plant host for a long time without causing any harm to host) streptomycetes can aid in their host plant growth by auxin production (Sadeghi et al. 2012).

2.4.4 Actinobacteria- as pathogen

Plant pathogenesis is a very complex multi step process. The pathogens have to infect the plant system, break the cell wall and win over the plant defense mechanism (Qin et al. 2011). Although, the number of actinobacterial plant pathogen is less however, sometimes they can cause a huge agricultural and economical loss by destroying huge amount of crops (Qin et al. 2011). *Clavibacter, Leifsonia, Rhodococcus* and *Streptomyces* are the most popular plant pathogenic actinomycetes. Most importantly, *Rhodococcus* and *Streptomyces* have ride range of host and thus are more deleterious than others. Most well known actinobacterial phyto-pathogenicity is the potato scab disease caused by *Streptomyces scabiei. S. turgidiscabies* and *S. acidiscabies* also causes scab disease. "Common" scab is caused by *S. scabiei* and *S. turgidiscabies* whereas; *S. acidiscabies* is responsible for "acid scab" (Loria et al. 1997). The stains of these two types of scab differ from each other in pigmentation pattern, pH resistant capacity and raffinose utilization capacity. Furthermore, *S. scabiei* and *S. turgidiscabies* can produce melanin however, *S. acidiscabies* is unable to produce the same (Nahar et al. 2018). Another form of scab called Russet scab is restricted to potato skin and quality of the crop. This disease is again of two types: American rudest scab and European russet scab. Though they are also caused by *Streptomyces* however, the causative strains are phytopathologically different from each other. Further, some streptomycetes (*S. flavovirens*) through their ligninocellulose degrding enzymes can harm the intact phloem cell wall of hosts and cause harm to them (Locci et al. 1994). *Rhodococcus fascians* have been reported to infect both monocot and dicot plants which

include economically important crops too (Cornelis et al. 2001). Upon infection, excess intracellular epiphytic growth through stomata leads to leaf deformation, fasciations, leafy galls and witched broom formation. Shoots are hyper-induced by dormant auxillary meristem activation and de novo meristem formation (Cornelis et al. 2001). Manipulation of host hormonal balance by the pathogen derived auxins and cytokines is the main reason behind the aforementioned symptoms (Cornelis et al. 2001).

Clavibacter michiganensis is the only reported plant pathogen of its genus (Daferera et al. 2003). All the five sub-strains of this species are potent plant pathogen colonizing in their host xylem. The severity of *C. michiganensis* infection and difficulty in controlling them has recognize this species as quarantine organism under the European Union Plant Health Legislation and laws of many other countries (Bentley et al. 2008). Ratoon stunting of sugarcane is a disease caused by *Leifsonia xyli* subsp. xyli (Monteiro-Vitorello et al. 2004). This is a systematic disease where plant growth is retarded. Putative fatty acid desaturation modifies the carotenoid biosynthesis pathway and produce Abscisic acid which inhibit the plant growth. One interesting fact regarding *Leifsonia* infection is, they are very much xylem limited (Young et al. 2006). Moreover, members of Mycobacterium, Corynebacterium , Propionebacterium and Tropherymae are popular as animal pathogen.

2.4.5 Actinobacteria- as a source of natural products

Actinomycin was the first discovered antibiotic produced from *Streptomyces antibioticus* in 1940 (Katz 1967). Later, streptothricin and streptomycin were also produced from *S. lavendulae* and *S. griseus* respectively (Handelsman et al. 1998). With advancement of research, actinobacteria proved itself as a great source and reservoir of natural products. Thus, this group has become important from the biotechnological perspective (Strobel et al. 2004). Aminoglycosides (neomycin, kanamycin, streptomycin), beta-lactum, anthracyclines, chloramphenicol, glycopeptides, lantibiotics, streptogramins, tetracyclins etc. are some other major antibiotics produced from actinomycetes. The antibiotic production capacity of each microbe also varies significantly. For instance,

some *Streptomyces* strains produce single antibiotic whereas, some are capable of producing a broad range of antibiotics (Strobel et al. 2003). Concurrently, actinomycetes are also popular as source of other secondary metabolites such as, herbicides, insectisides, immune-suppresors, antifungal agents etc. *S. aureus* S-3466 has been reported to biosynthesise a mixture of tetranactin and dinactin which are used for commercial production of macrotetrolides (Handelsman et al. 1998). This compound act as as insecticides and immuno-suparresive agents. Another example is Mildiomycin which act as an antifungal agent. Their primary sites of actions are locations of cell wall chitin synthesis, inositol biosynthesis as well as sites of protein and DNA synthesis (Handelsman et al. 1998). Similarly, validamycin, kasugamycin, polyoxinin B and D are also potent natural antifungal agents synthesised from different strains of *Streptomyces*. Validimycin A, produced from *S. hygroscopicus* var. limoneus is converted into validoxylamine A within fungal cell. This compound inhibits the breakdown of trehalose and thus suppresses the supply of carbohydrate to fungi and insects. Since, vertebrates do not depend upon the breakdown (hydrolysis) of disaccharide trehalose, the functionality of validamycin is biologically selective. Furthermore, Polyoxin B act as antifungal agent in friuts, vegetables and ornamental plants whereas, polyoxin D can control the growth of *Rhizoctonia solani* which causes therice sheath blight by inhibiting the microbial protein biosynthesis (Handelsman et al. 1998). Kasugamycin is active against rice blast (*Pyricularia oryzae cavara*) and *Pseudomonas* disease in several crops (Rossman et al. 1990).

Recent genomic analysis of Actinobacteria reported that, some secondary metabolite biosynthetic machineries regulated by selected genetic loci may become active under precise lab conditions which may further enhance the commercial importance of these microbes as a natural resource of biologically active compounds.

2.4.6 Probiotic Actinobacteria

Advancement of research has proved the vital role played by gut microbiome in various aspects of human health. *Bifidobacterium*, is one of the major constituent of gut microflora which is

ubiquitous in every human (Ishibashi et al. 1997). The population of these microbe changes with age. In childhood they are enormous and unstable whereas, in adulthood their population comparatively reduces and become stable. *Bifidobacterium longum*, *B. breve*, and *B. bifidum* predominate in infants. On contrary, *B. catenulatum, B. adolescentis* and *B. longum* are more prevalent in adulthood (Lau and Liong 2014). Several extrinsic factors are important for the initial colonization of this genus in intestine. Infants can also acquire this microbial population from mother through mother's vaginal tract, GIT, breast milk, placenta and amniotic fluid (Wei et al. 2012). Further, birth mode is another important factor in this case. Infant born vaginally are more abundant with bifidobacterial population than infant born by cesarean section. The gut microbiome may vary with the feeding regime too. For example, *Bifidobacterium breve, B. bifidum*, *B. longum* ssp. longum, and *B. longum* spp. infantis have been reported to be present in both breast- and formula-fed infants. *B. longum* ssp. infantis and *B. longum* ssp. longum are commonly associated with breast-fed infants and bottle-fed infants respectively (Ishibashi et al. 1997, Wei et al. 2012). Further, formula fed babies do not develop *B. adolescentis* in their adulthood. Thus, breast milk does have a bifidogenic effect since; selected glycans present in breast milk can be utilized by *Bifidobacterium* (Wei et al. 2012). This report was also supported by another finding that, breast milk of mother with mutation in Fucosyl transferase 2 gene (FUT2) delays the colonization of gut Bifidobacteria since, the mentioned mutation inhibit the conversion of fucose to glycans (Ishibashi et al. 1997).

Bifidobacterium composition in human intestine also fluctuates with different disease condition. Obesity, one of the major condition affecting children and adults has been shown to have correlation with alterations in gut microbiome. A relatively lower set of *Bifidobacterium* and higher set of Staphylococcus have been reported previously in obese children (Million et al. 2012). Those mothers who during their pregnancy period gain over weight supply lower amount of *Bifidobacterium* to their babies than that of healthy normal weight mother. Besides, patients with short or long term asthma also bear lower amount of

aforementioned strains.*B. adolescentis* has been found to be dominant in asthma patients. This strain is generally found in adulthood however, if a mother has allergy, the baby may get the strain through breast feeding. On contrary, reports are there suggesting the probiotic and anti allergic effect of *B. longum* in children (Ouwehand et al. 2001). Further, during irritable bowel syndrome, the population of *Bifidobacterium* also decreases. Thus, the variety and population of Bifidobacteria has a relation with different syndrome, suggesting their role in regulating that pathophysiological condition (Shirakawa et al. 2018). Moreover, their capacity to stimulate the immune system has recognized them as probiotics. However, this capacity has been found to be specie specific. Systematic and intestinal immunity is very much elevated by *B. bifidum, B. dentium,* and *B. longum*. Members of these genera are also used in treatment against allergies, celiac disease, obesity, diarrheas, colic, infections or necrotizing enterocolitis (Vazquez et al. 2010). The bile sale hydrolase activity of some selected strains reduces the blood cholesterol and lipoprotein cholesterol level hence, can also be useful in restoration of lipoprotein imbalance in children having dyslipidemia (Sharma et al. 2016). Concurrently, the probiotic strains also deal with some brain gut disorders like anxiety, depression, stress etc. Interestingly, Bifidobacteria is now being used in diagnosis of tumor since; they have the ability to colonize at tumor sites. They can also inhibit the growth of tumor by tumor specific T-cell responses, CD8+ T cell responses and dendritic function modulation (Kim et al. 2007). Thus, Bifidobacteria with their probiotic activities have become some of the major genera of Actinobacteria.

2.5 Extremotolerant Actinobacteria

Actionobacteria can be found both in normal as well as in extreme nutrient-deprived environments characterized by low or high temperatures, acidic/ alkaline pH, high radiation, salinity, low levels of available nutrients, and moisture (Zenova et al., 2011). The varied physiology as well as metabolic elasticity of extremotolerant actinobacteria enables them to endure unfavourable and hostile circumstances. Bull (2010) recorded the high abundance of actinobacteria under extreme conditions which broke the conventional concept of restricted

preponderance of actinobacteria in soil and fresh water environment. Several studies have been carried out on actinobacteria isolated from normal environments (neutral pH and temperature ranging 20–40°C). However, only a few reports are available to understand the diversity of actinobacteria present in extreme environments, their adaptation and ecological role. Poly-extremotolerant and poly- extremophiles actinobacteria are also capable to exist in environments with more than two extreme conditions. Poly-extremophiles are able to live in environments having multiple stresses (Gupta et al., 2014). Thermo-acidophilic, alkali-thermophilic, haloalkaliphilic, thermophilic radiotolerant, and thermo-alkalitolerant actinobacteria are some examples of Poly-extremophiles. Their occurrence has been recognized in different extreme geographical locations including the Arctic (Augustine et al., 2012) and Antarctic (Gousterova et al., 2014) regions, hot springs (Chitte and Dey, 2002), oceans (Raut et al., 2013), and deserts (Kurapova et al., 2012).

The extremophilic actinobacteria show numerous adaptive schemes for example- switching between two different metabolic modes (i.e., heterotrophy, autotrophy, and saprobes) antibiosis and fabrication of specific enzymes which may help them to survive under harsh conditions (high temperature, saline and alkaline). The thermo-tolerance is accredited to the presence of elevated hydrophobic and electrostatic interactions as well as disulfide bonds in the proteins of thermophilic strains (Ladenstein and Ren, 2006). They have definite unusual proteins called chaperones which help in refolding the proteins with partial denaturation (Singh et al., 2010). Several other proteins are also present which bind to DNA and avoid their denaturation at high temperatures. A thermophilic *Streptomyces* sp., obtained from desolated place was reported to produce autotrophic metabolic pathway enzymes like carbon monoxide dehydrogenase (CODH) (Gadkari et al., 1990). This enzyme facilitates the growth of microbes in nutrient deprived conditions via oxidizing the accessible inorganic compound such as carbon monoxide into CO_2 which can further fixed by the enzyme RuBisCO into microbial biomass by Calvin–Benson cycle (King and Weber, 2007). The chemolitho-autotroph thermophilic,

Acidithiomicrobium sp., isolated from geothermal environment, exploits sulphur as their energy source (Norris et al., 2011).The antibiosis is an additional strategy through which actinobacteria keep going by killing other microbial flora under nutrient restricted conditions. Alkaliphiles and acidophiles regulate H+ concentrations inside and outside the cell through proton pumps for maintaining their physiological pH inside (Kumar et al., 2011). Alkaliphiles have negatively charged cell wall polymers which can stabilize cell membrane by charge density reduction at the cell surface (Wiegel and Kevbrin, 2004). The adaptive strategy of halo-alkaliphiles embraces excess tolerance to salty environment by producing and accumulating higher amount of attuned solutes (Roberts, 2005) which can avoid desiccation through osmo-regulation. Additionally, they have Na^+/H^+ antiporter system to exclude undue salt from inside of the cell.

Some actinobacteria also have the capacity to live in various radioactive sites and tolerate extremely harmful radiations like UV and gamma rays. Three thermophiles of *Rubrobacter* species - *R. xylanophilus* (Ferreira et al., 1999), *R. radiotolerans* and *R. taiwanensis* (Chen et al., 2004) were reported to be radio-tolerant. The resistance machinery has not been sufficiently understood, but the whole genome analysis of *R. radiotolerans* RSPS-4 exposed the presence of genes responsible for encoding proteins those are involved in oxidative stress response, DNA repair system and biosynthetic pathways of companionable sugars (mannosyl glycerate and trehalose) playing a vital role in extenuating the damage caused via radiation (Egas et al., 2014). In current years, some more radio tolerant and alkali tolerant actinobacterial species such as *Microbacterium radiodurans* GIMN 1.002T (Zhang et al., 2010), *Microbacterium maritypicum* (Williams et al., 2007), *Cellulosimicrobium cellulans* UVP1 (Gabani et al., 2012), *Kocuria* sp. ASB107 (Asgarani et al., 2012) and *Kocuria rosea* strain MG2 (Gholami et al., 2015) have been reported. The two alkali tolerant *Kocuria* strains were. The *Kocuria* sp. ASB107 isolated from Ab-e-Siah radioactive spring of Iran is a psychrotrophic strain showing tolerance to ionizing radiation (upto 90% lethal doses) such as corona discharge and ultraviolet (256nm-UV). The *Kocuria rosea* strain MG2 was

revealed to tolerate the high quantity of damaging UV-C radiation. This strain can grow in a ample pH range (5–11 with optimum growth at pH 9.2) along with salt concentration (0– 15%). Cell-viability analysis on *Kocuria rosea* strain MG2 under multiple stresses was reported by Gholami et al. (2015). It was revealed that, after 28 days of incubation under desiccation condition, the *Kocuria* strain cells were viable and demonstrated high tolerance to the strong oxidant such as H_2O_2 (1–4%) and radiation. H_2O_2 is a well accepted antimicrobial compound damaging biological membranes through hydroxyl radical generation. Hence, the aforementioned *Kocuria* strains seem to display both non-enzymatic anti oxidant defense systems (carotenoids) and enzymatic (catalase and peroxidase) diminishing the consequence of radiation, strong oxidants and other stresses (Gholami et al., 2015).

The flexibility and adaptability of extremophilic/ extremotolerant actinobacteria award them a competitive benefit over other microbes. Besides helping them to endure extreme conditions, the physiology as well as metabolic elasticity also prompts them to create industrially costly compounds (Singh et al., 2013). The manufacture of biomolecules through extremophiles alleviates the risks of other bacterial contaminations, besides providing halo -tolerant, alkali-stable and thermo-stable compounds. Enzymes produced through these extremophilic/ extremotolerant actinobacteria are also useful under extreme environment, thus, making them appropriate candidates for industrial processes with harsh treatment methods.

2.5.1 Thermophilic and Thermotolerant Actinobacteria

Thermophilic actinobacteria flourish at comparatively elevated temperature ranging from 40 - 80°C (Tortora et al., 2007). They are generally found in moldy silage (Corbaz et al., 1963), self -heating residues of plant, sugar cane bagasse, cereal grains (Suihko et al., 2006), compost heaps, decaying vegetable materials (Shivlata and Tulasi 2015).These are 2 types: stringently thermophilic which can grow in the temperature range between 37 and 65°Cwith optimum proliferation 55–60°Cand reasonably thermophilic which thrive at 28–60°C and need 45–55°C for their optimum growth (Jiang and Xu, 1993). Another collection called thermo-tolerant actinobacteria is

able to endure upto 50°C temperature (Lengeler et al., 1999).

Thus, actinobacteria being a diverse group of microbes are present in all probable habitats and are important from medicinal, ecological and environmental aspects. In spite of all diversities, there remain some unities among these organisms in both genomics and proteomic level. Those signatures, if identified, will definitely help us to fill up the pit holes regarding our understanding about actinobacteria. In this study, we are going to utilize the power of bioinformatics to analyze the massive-data generated from a mass of actinobacteria to obtain an overall "birds'-eye" view of these microbes.

2.6 Bioinformatics as tool to explore the complexities microbial world:

2.6.1 Dawn of bioinformatics:

It was in the year 1953 when James Watson and Francis Crick proposed the twisted-ladder (double helix) structure of deoxyribonucleic acid (DNA). This particular event forever changed the history of biological science and gave rise to modern molecular biology. In short order, their discovery yielded ground-breaking insights into the genetic code and protein synthesis. During the 1970s and 1980s, it helped to produce new and powerful scientific techniques, specifically recombinant DNA research, genetic engineering, rapid gene sequencing, etc. Around the same time, the term "bioinformatics" was coined by Ben Hesper and Paulien Hogeweg (Hesper and Hogeweg, 1970). Bioinformatics is by nature a cross-disciplinary field that began in the 1970s with the efforts of Margaret O. Dayhoff, Walter M. Fitch, Russell F. Doolittle and others and has gradually matured into a fully developed discipline. Initially, it was referred as 'the study of information processes in biotic systems' (Hesper and Hogeweg, 1970). However, its primary use since at least the late 1980s has been to describe the application of computer science and information sciences to the analyses of biological data, particularly in the areas of genomics involving large-scale DNA sequencing (Luscombe et al., 2001). The arrival of the INTERNET is another important milestone in the development of bioinformatics as a full-fledged discipline. This discipline represents the convergence of genomics, biotechnology and information technology, and encompasses analysis and

interpretation of data, modeling of biological phenomena, and development of algorithms and statistics (Fenstermacher, 2005). The need for bioinformatics was further accelerated when the Human Genome Project (HGP) was launched in 1990. The aim of the project was to sequence the entire human genome. Information gleaned from the HGP is not very useful until the huge data is managed and interpreted in a proper way by the computational tools leading to the materialization of bioinformatics. The success of HGP opened the flood-gates for other genome sequencing projects. Gradually genome sequences of mouse, rat, worms, yeast and plants like rice, Arabidopsis were completed. The publication of huge amount of sequence data were greatly supported by development of high end computers, smart computational tools for large-scale annotation, functional classification of the proteins (Searls, 2000) and development of specific databases (Birney et al., 2002). Availability of complete genome sequences of different organisms lead to the development of public repositories of gene data like GenBank (Benson et al., 2000), EMBL (Baker et al., 2000), DDBJ (Okayama et al., 1998), Protein DataBank (PDB) (Bernstein et al., 1977) and several others. After the formation of the databases, tools became available to execute various analyses. Two programs, which greatly facilitated the similarity search, were FASTA (Pearson and Lipman, 1988) and BLAST (Altschul et al., 1990). Many programs have been further developed since then. Accessibility of free and open source software has taken bioinformatics and its application to all -together new heights.

2.6.2 Bioinformatics based platforms suited for microbial research:

Several 'omics' based research disciplines i.e., genomics (for genome data), proteomics (for protein data), transcriptomics (for gene transcription data), etc., have emerged with the enormous advancement of bioinformatics. The 'cause and effect' relationships of biological systems are being frequently employed by mathematical models and computational simulations for proper elucidation of biological complexities. Proper blend of knowledge-driven computational simulations and data-oriented bioinformatics holds the key in apt realization of the riddles of host-microbe interactions. Meaningful

analysis of high-throughput sequencing data promises to unravel the enigma of the complex associations. Specific biological databases, providing access to annotated genomic data of bacterial populations, like IMG/M (Markowitz et al., 2014), SEED (Overbeek et al., 2005) and Greengenes (DeSantis et al., 2006) have opened up revenues to extract significant information. Annotation pipelines like RAST (Aziz et al., 2008) and MG-RAST (Meyer et al., 2008) have been instrumental for the wholesome purpose of characterizing bacterial masses. Systems Biology Markup Language (SBML) (Hucka et al., 2003) and BioModels repository (Le Novere et al., 2006) have been very useful for proper standardization of biochemical reaction models. Various bioinformatics based platforms suited for genomic investigations are as follows:

2.6.2.1 Marker Gene Profiling

Marker gene profiling (also known as gene amplicon sequencing) involves extraction of DNA from different samples. DNA extraction is followed by amplification using primers specifically designed against ribosomal RNA genes. Initially, 454 pyrosequencing technology was pretty popular for the chore but nowadays, it has been replaced by Illumina sequencing which has been reported to produce equivalent results with a better coverage (Luo et al., 2012). QIIME (Navas-Molina et al., 2013) and MEGAN (Huson et al., 2011) have been efficient pipelines for analyzing data extracted via marker gene profiling. PICRUSt effectively predicts the functional potentials of host associated microbes by analyzing the metagenome from the marker gene data (Langille et al., 2013). Marker gene profiling does not involve high cost and thus, aptly suits the purpose of large projects concerned with data extraction. Pathway enrichment followed by metabolic reconstruction has been suggested to be fruitful pertaining to proper analysis of metagenomic data (Abubucker et al., 2012; Sharon et al., 2011). Robust machine learning algorithms have been suggested to be useful for proper predictive analysis of data produced by marker gene profiling (Nakano et al., 2014; Statnikov et al., 2013). The major limitation of the technique is that it largely depends on pre-existing curated rRNA databases for taxonomic profiling which highlights the impotence of the method to

successfully characterize novel bacterial species present in the samples (Dicksved, 2008).

2.6.2.2 Metagenomics:

Metagenomics is the branch of genomics that performs investigations by direct extraction and cloning of DNA from assembly of organisms (Handelsman, 2004). Comprehensive sequencing is necessary to pursue metagenomics based investigations (Fig 2.1). It is taken into presumption that human gut microbiome contains a large number of uncultured species and therefore, thorough sequencing is mandatory to proceed with metagenomics of gut samples (Emerson et al., 2012; Hess et al., 2011; Narasingarao et al., 2012). Sophisticated genome assembly protocols (Bashir et al., 2012; Goldberg et al., 2006) and efficient single-cell sequencing technologies (Lasken, 2012; Shapiro et al., 2013) have been suggested to aid proper metagenomic analysis. Due to extensive focus on host associated microbiome and availability of numerous bacterial genomes, assembly -free methods tend to be useful in thorough scrutiny of metagenomic samples (Carr et al., 2013; Luo and Moran, 2013). Rapid progress of next-generation sequencing (NGS) technologies has been a boost for the domain of metagenomics. NGS based techniques allow in-depth profiling and analysis of microbiome without the limitations of selection bias and constraints, associated with cultivation methods. NGS-based methods rely greatly on sophisticated bioinformatics based tools, regularly updated data repositories and functional know-how.

The main motif of bacterial metagenomic analysis has been proficient characterization of and profiling of the microbial members framing the diverse microflora. Similarity based alignment tools like BLAST serve the preliminary purpose of finding out regions of similarity between the raw or assembled sequences against a reference database and thus, provide an initial clue about the functionalities of the query samples. MG-RAST pipeline employs the M5nr database that is a large repository of non-redundant protein sequences from multiple sources (Weinstock, 2012; Wilke et al., 2012). QIIME pipeline facilitates taxonomic classifications. Functional and metabolic databases like KEGG Orthology (Kanehisa and Goto, 2000) or SEED subsystems (Overbeek et al.,

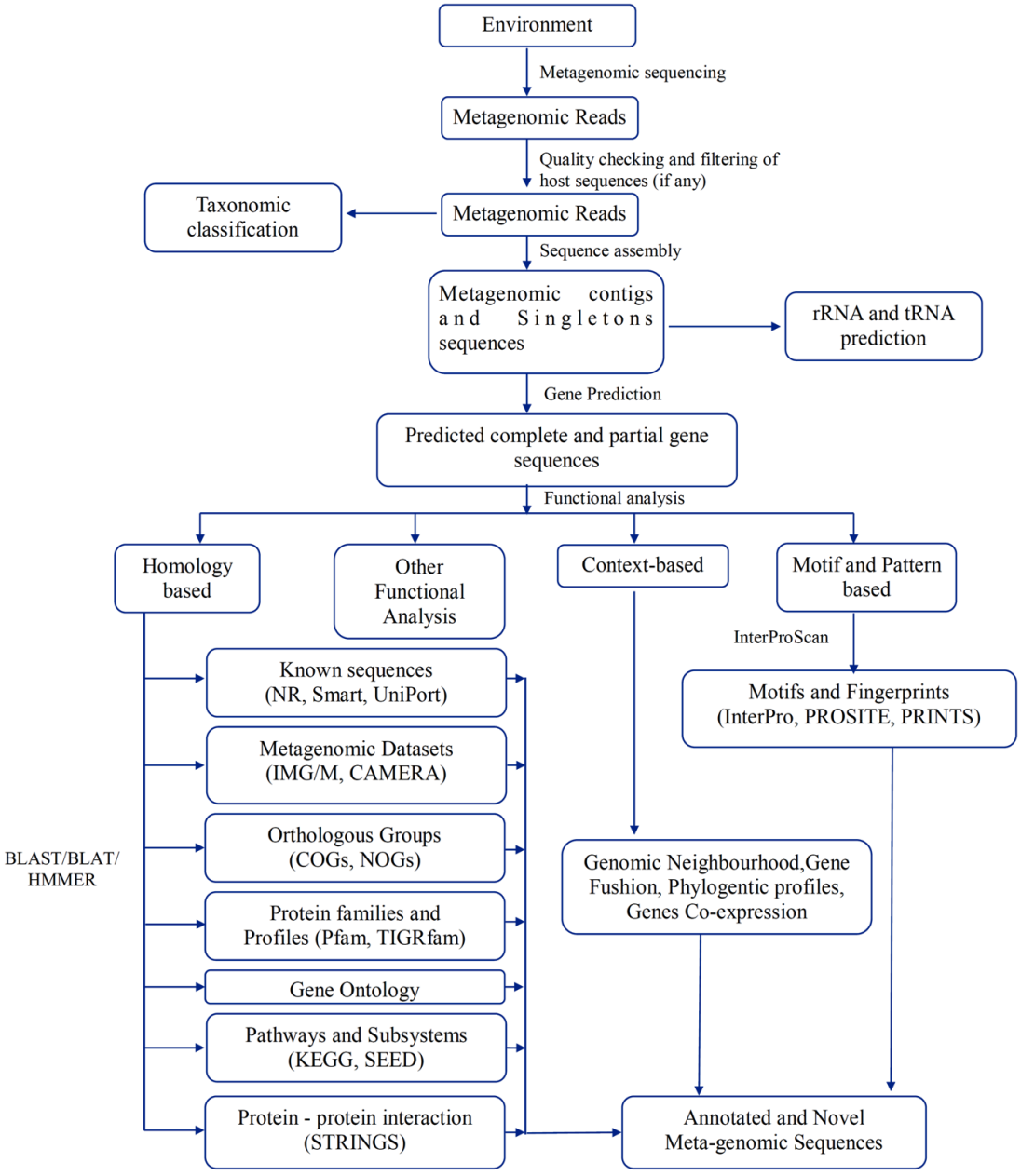

Fig 2.1 Schematic diagram for Metagenomics data analysis. The scheme was adopted from Prakash, T. and Taylor, T.D., 2012. Functional assignment of metagenomic data: challenges and applications. Briefings in bioinformatics, 13(6), pp.711-727.

2005) assist investigations pertaining to functionalities. Sophisticated statistical techniques are also frequently employed for extensive functional analysis (Kristiansson et al., 2009). Metabolic modeling tends to be useful in apt execution of function oriented studies (Abubucker et al., 2012; Jiao et al., 2013; Levy and Borenstein, 2013).

2.6.2.3 Metatranscriptomics:

Metatranscriptomics involves

extraction of RNA from samples of interest and is much more complex than DNA extraction technologies. The major concern of metatranscriptomics is to achieve good quality sequence and sufficient yield. Fast alignment tools like bowtie (Langmead et al., 2009) and SSAHA (Ning et al., 2001) find applications in proper alignment of transcriptomic data, retrieved from the host-microbiota sample, with a set of representative bacterial genomes (McNulty et al., 2011; Turnbaugh et al., 2010; Xiong et al., 2012). Simple BLAST has also been employed for the purpose (Gosalbes et al., 2011; Xiong et al., 2012) but the task requires extensive and rigorous computational support to align millions of sequence reads. Current pipelines use variations of this basic approach (Leimena et al., 2013; Xu et al., 2014). Metatranscriptomic data promise to be handy in exploring the intricacies of gut microbiota and provide vivid depiction of the host-microbiota interactome (Westermann et al., 2012; Xu et al., 2014). However, retrieval of sufficient bacterial RNA from combined host-microbiota sample has been a daunting task as host RNA prevails over bacterial RNA due to low biomass of the microbial members. Another crucial step is the removal of ribosomal RNA from bacterial RNA in order to avoid misinterpretations while analyzing metatranscriptomic data.

2.6.2.4 Metabolomics:

Metabolomes refer to the complete set of metabolites present in a cell/ organism or a community of organisms (Jordan et al., 2009). Research pertaining to metabolome of human gut microbiota has been an exciting challenge these days. Metabolomes have been suggested to be crucial players of proper interactions and cross -talks between gut associated microbes and human host (Marcobal et al., 2013; Nicholson and Lindon, 2008). Extensive profiling of metabolomes promise to confer large body of information regarding the complex interactions and also the health benefits and hazards associated with the concerned microbes of human gut (Larsen and Dai, 2015).

2.6.2.5 Computational Modeling and Simulation:

Availability of complete genome sequences has made it feasible to execute computational and metabolic modeling of concerned microorganisms (Oberhardt et al., 2009; Thiele and Palsson, 2010). Huge

data surging out of the various metagenomic projects also provide ample scope for apt simulation and metabolic modeling (Henry et al., 2010). Computational simulations do employ simplified techniques based on abstractions. Computational models not only prove fruitful for a single organism but also can be employed efficiently for a community of organisms displaying significant interactions (Borenstein, 2012). This technique also finds its applications in supra-organism where there is significant assemblage of metabolic pathway associated genes reflecting all involved species (Borenstein, 2012).