Evolution of *M. tuberculosis* genomes: Lessons from comparative genomics analysis

Sarbashis Das, Anchal Vishnoi, Rahul Roy* and Alok Bhattacharya,

School of Computational & Integrative Sciences, Jawaharlal Nehru University, and *Indian statistical Institute, New Delhi

A number of genomes have been completely sequenced and the list is increasing everyday. Particularly resequencing of different strains using next generation sequencing methods has allowed creation of large organism specific data for strain evolution analysis. Comparative analysis of these fully sequenced genomes, particularly closely related organisms, is likely to help in establishing sequence-phenotype relationships. Tuberculosis is a major problem through out the world killing millions every year. There has been a surge in the disease burden with emergence of disease burden in the causative organism *M. tuberculosis*. A number of molecular processes contribute to genomic diversity. The major differences that contribute to diversity are insertion and deletion, single nucleotide polymorphism (SNP), duplication, repeat expansion and recombination which can change order of the genes. All these can contribute to phenotypic divergence depending on the situation. Therefore it is important to identify all the changes that are present between any two genomes for detailed biological studies. We have developed new algorithms and computation pipelines for automated discovery of genomic variants from fully sequenced and from next generation sequencing reads of microbial genomes. We have identified single nucleotide variations in *M. tuberculosis* using genome sequences from 25 strains from throughout the world. Using different statistical tools we have identified a pattern in genomic variations. We hope that our studies will help to understand the relation between phenotype and genotype.