

## Poster A-21

### Anchor-based whole genome divergence and phylogeny: Analysis of Mycobacterial genomes



#### Authors:

Anchal Vishnoi (*Research Scholar*)

Rahul Roy (*Professor*)

Alok Bhattacharya (*Professor*)

**Short Abstract:** Comparative genome analysis is a powerful approach to measure the diversity among the genomes, their evolution and pathogenesis. We had applied a sampling approach based on local and global information of the whole genome to identify the course of evolution and divergent regions among the strains of Mycobacteria.

#### Long Abstract:

Comparative genome analysis emerges out as a powerful approach to measure the diversity among the genomes. A number of experimental and computational approaches based on sequence divergence, single nucleotide polymorphisms (SNP), repetitive sequence based polymorphisms, variable transposon insertions, recombination events had been used in past to decipher the biological features of the closely related genomes [1,2,3,4]. All these methods help in understanding the evolution, identification of regions of divergence, pathogenesis. One of the major problem of understanding the host- pathogen relationship is to explain the clinical features ranging from asymptomatic infection to different forms of invasive diseases. These differences can be attributed to genetic differences between the various strain. The acquisition of genes because of horizontal gene transfer is a major cause for attaining virulent genes for example, in *Escherichia coli*, the pathogenic strain *E.coli* O157:H7 acquired 13.4 Mb of strain specific sequences [5], also in *Helicobacter pylori* pathogenicity islands are attained by similar mechanism which are thought to be responsible for severe form of ulcers [6]. Among mycobacteria SNPs, insertion elements and genomic deletions are thought to be important for virulence in different strains and species [7]. There is no evidence of a set of toxin genes that can be directly associated with the virulence as described earlier. Mostowy et al [8] had shown that sequence polymorphisms can be ascribed to a pattern of deletion and the genes affected by these deletions may have a major role in virulence. These studies clearly show that different genomes evolve by different mechanisms and the changes caused by these genomic events influence their biological properties. Therefore, a study of the phylogenomic analysis among different strains and species of an organism will lead to an understanding of the evolutionary mechanisms leading to pathogenicity.

There are various methods available for whole genome tree construction. These methods are based on different genome properties such as statistical properties of sequences, similarity measure, gene content and gene order [9]. However, these studies do not throw any light on the virulence mechanisms which can be correlated with the pathogenesis. Thus, a method which can identify the divergent regions as well as determine the cause of emergence of these regions taking into account the evolutionary relationship between the closely related genomes is needed. Our method tries to fulfill this aim, it is based on sampling method which

uses both the local and global information for the inference of the evolutionary relationship between the genomes. The clustering approach is used to identify the regions of divergences. We have determined the evolutionary relationship using various genomic features and whole genome sequences. We have also identifies the regions of divergence between the two genomes and found out the mechanisms leading to these regions. We have applied the method to analyze the set of genomic sequences from various strains of Mycobacteria.

1. Tomb J. F. et al.1997. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature*. 388:539-547.
2. Ferretti J.J. et al 2001. Complete genome sequence of an M1 strain of *Streptococcus pyogenes*. *PNAS* 98(8). 4658-4663.
3. Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A., Shiba T., Ogasawara N., Hattori M., Kuhara S and Hayashi H. 2002. Complete genome sequence of *Clostridium perfringens* an anaerobic flesh eater. *PNAS* 99(2). 996-1001.
4. Beres S.B. Et al.2002. Genome sequence of serotype M3 strain of group A *Streptococcus* : Phage encoded toxins, the high virulence phenotype and clone emergence. *PNAS*. 99(15). 10078-10083.
5. T Hayashi, K Makino, M Ohnishi, K Kurokawa, K Ishii, K Yokoyama, C. G. Han, E Ohtsubo, K Nakayama et al 2001. Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. *DNA Res*. 8:11-22.
6. N Salama, K Guillemin, Daniel Mc, G Sherlock, L Tompkins, S Falkow.2000. A whole genome microarray reveals genetic diversity among *Helicobacter pylori* strains. *Proc Natl Acad Sci USA*. 97:14668-14673.
7. Fleischmann RD, Alland D, Eisen JA, Carpenter L, White O, Peterson J, DeBoy R, Dodson R, Gwinn M, Haft D {et al}: Whole genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains. {*J. Bacteriol.*} 2002, 184(19):5479-90.
8. S Mostowy, J. Inwald , S. Gordon ,C. Martin , R. Warren, K. Kermer , D. Cousins and M.A. Behr. 2005. Revisiting the evolution of *Mycobacteria bovis*. *Journal of Bacteriology*. 187(18): 6386-6395.
9. B. Snel, M.A. Huynen and B.E. Dutilh. 2005. Genome Trees and the Nature of Genome Evolution. *Annual Review Microbiology*. 59: 191-209.