EVALUATING THE PHYLOGENETIC ACCURACY OF SPOLIGOTYPING FOR *M. TUBERCULOSIS* EPIDEMIOLOGY

Type of project*:

Bachelor End/Honors Project
Master End Project (thesis or internship)

Student background:** Bachelor/Master CS/LST/NB

Background

One of the most widely used fingerprinting methods for M. tuberculosis is the CRISPR-based system referred to as Spoligotyping. This is due to its low cost and ease of implementation in low resource settings. A whole genome SNP-based approach to assigning M. tuberculosis strains to lineages and sub-lineages also exists (Coll et al. 2014). It is currently unknown how well these two approaches to typing M. tuberculosis strains correlate, although previous papers have suggested that Spoligotyping suffers greatly from pattern convergence (the same pattern arising independently in different strains).

The goal of this project is to create a map of Spoligotype pattern to lineage numbering for the entire M. tuberculosis complex and assess the usefulness of Spoligotyping for robustly genotyping strains.

Aims

A map of Spoligotyping to the high-level lineage numbering (1-6) exists (Comas et al. 2009) but has not been updated to include new lineages or any of the sub-lineages and is based on a small number of samples. Using the vast number of publicly available genomes (5000+) in addition to large datasets from endemic countries (e.g. Gambia, Bangladesh), this project aims to create a new map with increased resolution, preferably implemented in a way that can be easily repeated when new genomes are sequenced. Additionally, phylogenetic metrics will be used to find where Spoligotype and Lineage assignments robustly agree and where they do not, indicating if and when Spoligotype patterns can reliably be used for M. tuberculosis epidemiology.

Expected outcome: A map of Spoligotyping patterns to sub-lineage and an assessment of how well Spoligotype captures genetic relatedness.

Suggested literature

Coll et al. (2014) A robust SNP barcode for typing Mycobacterium tuberculosis complex strains. Nature Communications

Comas et al. (2009) Genotyping of Genetically Monomorphic Bacteria: DNA Sequencing in *Mycobacterium tuberculosis* Highlights the Limitations of Current Methodologies. PLOS1

Responsible supervisor: Thomas Abeel – t.abeel@tudelft.nl - EEMCS Pattern Recognition and Bioinformatics

Partners:

- Dr. Conor Meehan (Institute of Tropical Medicine, Antwerp, Belgium) – daily supervisor
- Dr. Iñaki Comas (Instituto de Biomedicina, IBV-CSIC, Valencia, Spain)

---

*Types of project: Bachelor seminar (TI3706): 5 ECTS literature review course // Research Assignment (IN5010): 15 ECTS bioinformatics literature review // BEP: 10-15 ECTS Bachelor End Project // MEP: 30-60 ECTS Master End (thesis) Project // internship: 3 month no credit project. The type of project you are completing will impact the scope and depth that you will be expected to accomplish.

** Student background: The Delft Bioinformatics Lab serves a broad student community with a variety of projects. The background mentioned here is a suggestion and not a restriction. We will adapt the scope and focus of the project to connect well with your expertise and program requirements.