AUTOMATED VALIDATION *M. TUBERCULOSIS* WHOLE GENOME SEQUENCING PIPELINES

Type of project*:

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

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

Background

Whole genome sequencing (WGS) of Mycobacterium tuberculosis has rapidly evolved from a research tool to integration into diagnosis of drug resistant tuberculosis, patient management, and use of phylogenetics for public health surveillance. There is however a risk that, in the absence of a consensus and standards, the widespread use of WGS technology may result in data and processes that lack harmonisation, comparability and validation. Such harmonisation requires current WGS pipelines to be validated and compared before they can be implemented in a clinical setting.

The goal of this project is to create a tool that can automatically generate short read datasets containing specific whole genome sequence features for validation of WGS pipelines for M. tuberculosis. These datasets will then be used to validate and compare leading WGS pipelines for M. tuberculosis.

Aims

Specific genomic features are required to be detected by WGS pipelines before they can be considered ready for clinical use. Such features include detection of drug resistance related mutations, strain typing, distinguishing epidemiologically linked isolates etc. A program will be created where Illumina read sets will be generated which contain different mock isolates with varying features. These datasets will then be used to test varying pipelines and settings to determine their accuracy and downfalls.

The expected outcome is a tool for automatic generation of mock read sets containing specific genomic features and a comparison of current leading WGS pipelines for M. tuberculosis assembly.

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.