Rijksinstituut voor Volksgezondheid
en Milieu
Ministerie van Volksgezondheid,
Welzijn en Sport

Workshop

"Next generation sequencing and bioinformatics tools for *Mycobacterium tuberculosis* drug resistance detection and epidemiological analysis"

St. Petersburg, Russia, 9-10 September 2019

Co-organised by
National Institute for Public Health and the Environment, RIVM (Netherlands)
& St. Petersburg Pasteur Institute (Russia)

Venue: Conference Hall, St. Petersburg Pasteur Institute, 14 Mira street, St. Petersburg, Russia

Language: English.

Registration. Free. Please send to Igor Mokrousov imokrousov@mail.ru: your name and surname, position and affiliation, and email.
Access to lectures, practical workshop and coffee breaks: free for registered participants.

Monday, 9 September 2019

Co-chairs: Dick van Soolingen & Igor Mokrousov

10:00-10:20 **Opening of the Workshop.** Dick van Soolingen, Igor Mokrousov

10:20-11:00 **Dick van Soolingen**, *Professor, Tuberculosis Reference Laboratory, National Institute for Public Health and the Environment, RIVM and Professor at Radboud University (Netherlands).*

Utility of Whole Genome Sequencing of *Mycobacterium tuberculosis* in the daily practice of a tuberculosis laboratory

Whole Genome Sequencing of *Mycobacterium tuberculosis* has improved the diagnosis of tuberculosis significantly regarding (sub)species identification, drug susceptibility testing and epidemiological typing. After an extended validation in the international context this technique is now ready for broad implementation to replace multiple other laboratory methods.

11:00-11:30 Coffee break

11:30-12:00 **Richard Anthony**, *PhD, Tuberculosis Reference Laboratory, National Institute for Public Health and the Environment, RIVM (Netherlands).*

RIVM database: concept, practical problems, advantages and limitations of the structure

For the past two years the RIVM has developed and hosted a database of MDR-TB detected in Europe as a member of the EuSeqMyTB consortium initiated by the ECDC. The database allows genome sequence data from any MDT-TB cultured in Europe to be compared to other isolates providing the possibility to detect successful clones present in multiple countries.

12:00-12:30 **Han de Neeling**, *PhD, Tuberculosis Reference Laboratory, National Institute for Public Health and the Environment, RIVM (Netherlands).*

Bioinformatics analysis of *Mycobacterium tuberculosis* at RIVM

Our TB reference lab at the RIVM analyses app. 1000 M. tuberculosis complex isolates per year by Illumina sequencing. Reads are mapped to H37Rv reference genome; SNPs and Indels are called using Breseq. Output is read into a MySQL database. Genetic distances (subspecies, clades, clusters) and resistance mutations are determined using R statistics. Results are presented via a web interface based on Linux, Apache, MySQL and PHP. A European database has been set up for MDR-TB isolates from 22 EU countries.

12:30-14:00 Lunch

Co-chairs: Egor Shitikov & Richard Anthony

14:00-14:30 **Danila Zimenkov**, *PhD, Engelhardt Institute of Molecular Biology (Moscow, Russia)*

Minor genetic determinants of resistance revealed by NGS data

Kanamycin, amikacin, and capreomycin resistance of *M. tuberculosis* are caused by mutations in the *rrs, eis, tlyA*, and *whiB7* genes. The next-generation sequence analysis for resistant isolates, that have no mutations in consensus loci, was performed. Possible resistance-associated mutations are located in elongation factor EF-G, phosphotransferase Aph, hypothetical Rv0147, and aspartate aminotransferase AspC.

14:30-15:00 **Viacheslav Sinkov**, *PhD, Scientific Centre for Family Health and Human Reproduction Problems (Irkutsk, Russia)*

Practical use of BSATool (Bacterial SNP Annotation Tool) in routine analysis of mycobacterial genomes from clinical specimens

BSATool is a cross-platform program that works with vcf (variable call format) files, allows you to annotate genomic positions, search for SNPs from specified lists, generate concatenated nucleotide and amino acid sequences, calculate the Dn / Ds value for all genes in the genome (Nei-Gojobori (1986)) method). The main features of the program will be demonstrated on the example of the analysis of genome-wide sequences to search for fragments of mycobacterial DNA extracted with urine in patients with tuberculosis.

15:00-15:30 Coffee break

15:30-16:00 **Egor Shitikov**, *PhD*, **Andrei Guliaev**, *PhD, Federal Research and Clinical Centre of Physical-Chemical Medicine* (*Moscow, Russia)*

Personal experience of comparative genomics analysis of *Mycobacterium tuberculosis* Beijing B0/W148 cluster

Mycobacterium tuberculosis Beijing B0/W148 is one of the most widely distributed clusters in the Russian Federation. Members of the cluster represent a separate group on the phylogenetic tree and carry 59 specific polymorphisms, which partly explain the high virulence. Additionally, all strains possess large chromosomal rearrangements in the genome.

16:00-16:30 **Igor Mokrousov,** *PhD, DSc, St. Petersburg Pasteur Institute (St. Petersburg, Russia)*

Beyond Linux and command line. Personal experience with some on/offline tools for *M. tuberculosis* NGS data analysis

I will briefly share my experience with some of available (and supposed to be user-friendly) online tools and resources and one offline package that I use for different specific tasks of NGS data analysis, and will discuss their useful and inconvenient features.

16:30-17:00 Discussion

Tuesday, 10 September 2019

Co-chairs: Han de Neeling & Danila Zimenkov

10:00-10:30 **Jeroen Laros**, *PhD, Tuberculosis Reference Laboratory, National Institute for Public Health and the Environment, RIVM (Netherlands).*

Infrastructure for large-scale NGS data analysis.

Large scale analysis of Next Generation Sequencing (NGS) data requires a formidable computational- and data management infrastructure. Furthermore, stable pipelines, controlled from an automation system with connections to Laboratory Information Management Systems (LIMS) and other information systems is needed to ensure reproducible analysis. In this lecture we will cover the components and design of such a system and we will discuss some of the critical implementation details.

Coffee break (30 min)

Practical work: implementation of the RIVM pipeline for *M. tuberculosis* NGS data analysis **(Jeroen Laros, Han de Neeling, Richard Anthony, Dick van Soolingen)**

General discussion and joint data analysis (all interested participants).

Lunch (90 min)

Closing remarks