Spoligotyping of *Mycobacterium tuberculosis* – comparing *in vitro* and *in silico* strategies

Zofia Bakuła[1], Mikołaj Dziurzyński[2], Przemysław Decewicz[2], Edita Vasiliauskienė[3,4,5], Laima Vasiliauskaitė[3,4,5], Daiva Bakonytė[3], Katarzyna Roeske[1], Petras Stakėnas[3], and Tomasz Jagielski[1]

[1]Department of Medical Microbiology, Institute of Microbiology, Faculty of Biology, University of Warsaw, Poland
[2]Department of Environmental Microbiology and Biotechnology, Institute of Microbiology, Faculty of Biology, University of Warsaw, Poland
[3]Department of Immunology and Cell Biology, Institute of Biotechnology, Life Sciences Center, Vilnius University, Vilnius, Lithuania
[4]Department of Physiology, Biochemistry, Microbiology and Laboratory Medicine, Institute of Biomedical Sciences, Vilnius University, Vilnius, Lithuania
[5]Centre of Laboratory Medicine, Tuberculosis Laboratory, Vilnius University Hospital Santaros klinikos, Vilnius, Lithuania

The study was financed by the National Science Center «DAINA» Programme (2017/27/L/NZ6/03279) and by Research Council of Lithuania «DAINA» Programme (S-LL-18-4)

Background: Since the 1990s, when molecular techniques became easily accessible to the mycobacteriologists, different tools for better understanding the epidemiology of tuberculosis (TB) have been developed. One of the most widely used methods for exploring the genetic diversity of *Mycobacterium tuberculosis* is spoligotyping.
The objective of this study was to compare the spoligotyping results for *M. tuberculosis* clinical isolates, produced using *in vitro* and *in silico* approaches.

Methods: The study included 118 *M. tuberculosis* (58 MDR and 60 DS) isolates, recovered from as many patients from Poland (*n*=58) and Lithuania (*n*=60) between 2018 and 2019. Genomic DNA was extracted using PureLink Genomic DNA Mini Kit (Thremo Fisher Scientific, USA) or using a modified cetyltrimethylammonium bromide method. Spoligotyping *in vitro* was performed with a commercially available kit (Mapmygenome India Ltd., India), as per manufacturer's instructions. Spoligotype shared types (ST) and phylogenetic clades were assigned according to the SITVIT2 database (http://www.pasteur-guadeloupe.fr:8081/SITVIT2/). Whole genome sequencing was done with Illumina NovaSeq 6000 sequencer in 2x150 bp paired-end mode. Phylogenetic clades of *M. tuberculosis* were assigned *in silico*, using three different spoligotyping tools, i.e. (i) *SpoTyping* (https://github.com/xiaeryu/SpoTyping-v2.0); (ii) *SpolPred* www.pathogenseq.org/spolpred and (iii) *lorikeet* (http://genomeview.org/jenkins/lorikeet/).

Results: Upon *in vitro* spoligotyping, the isolates produced 38 different profiles split into 14 clusters (*n*=94, 79.7%, 2-33 isolates per cluster) and 24 (20.3%) unique patterns. Most isolates belonged to the Beijing family (*n*=40; 33.9%), followed by T (*n*=26; 22%), Ural (*n*=15; 12.7%), Haarlem (*n*=11; 9.3%), and LAM (*n*=10; 8.5%) clades. Sixteen (13.6%) isolates were designated as Unknown/Not defined. Among MDR *M. tuberculosis* isolates, the most abundant were Beijing (*n*=36; 62.1%) and Ural (*n*=11; 19%) lineages.
Spoligotypes inferred from the WGS data were congruent with *in vitro* generated profiles in 81.3%, when *lorikeet* and *SpoTyping* tools were applied, or 74.6% if *SpolPred* was used. Thus, either 22

isolates (12 spoligotypes) or 30 isolates (18 spoligotypes) were differently assigned, as compared with *in vitro* profiling.

Conclusions: Given a relatively high (*ca*. 20%) discordance of the *in vitro* and *in silico* spoligotyping results, we advise to perform this genotyping as a conventional, PCR-reverse hybridization assay, at least unless more accurate tools are not available.
The spoligotype-based structure of the MDR *M. tuberculosis* population was conspicuously compact, since more than 80% of the isolates belonged to either of two lineages (Beijing and Ural). Among DS isolates, the T lineage predominated, comprising close to a third of the isolates.