Molecular characterization of *Mycobacterium tuberculosis* isolates from pulmonary tuberculosis patients in Sri Lanka

Sri Lanka is a South Asian country with moderate tuberculosis (TB) burden. Although genetic diversity of *Mycobacterium tuberculosis* (MTB) population will provide important data to monitor and underpin the Sri Lankan TB control programme, the molecular epidemiology of MTB is poorly explored.

In first chapter, I focused on identifying circulating lineages/sub lineages of MTB and their transmission patterns. DNA from 85 isolates of MTB collected during 2012 and 2013 from new pulmonary tuberculosis patients in Kandy District, Sri Lanka and analyzed by spoligotyping, large sequence polymorphism (LSP), mycobacterial interspersed repetitive unit-variable number tandem repeat (MIRU-VNTR) typing and drug resistance-associated gene sequencing. The predominant lineage was lineage 4 (Euro-American, 46.1%), followed by lineage 1 (Indo-Oceanic, 29.6%), lineage 2 (East-Asian, 23.6%) and lineage 3(Central-Asian, 1.2%). Among 26 spoligotype patterns, eight were undesignated or new types and seven of these belonged to lineage 4. Undesignated lineage 4/SIT 124 (n=2/8) and SIT 3234 (n=8/8) clustered together based on 24-locus MIRU-VNTR typing. The dominant sublineage was Beijing/SIT 1 (n=19), with isoniazid resistance *katG* G944C mutation (Ser315Thr) detected in two of them.

As the predominant lineage of MTB in Kandy, Sri Lanka was lineage 4, but not...
lineage 1 as expected, I suggest that lineage 4 may have been introduced to Sri Lanka during colonial period and undergone evolution to be adapted to the local hosts. Therefore, the chapter II was performed to detect the genomic variations in MTB lineage 4 and to identify the clonality and micro diversity of SIT 3234 in Kandy, Sri Lanka. Genomic DNA of 20 isolates of lineage 4 were sequenced using Illumina MiSeq sequencer. The MTB H37Rv genome (NC_000962.3) was used as the reference genome in analysis. Six sublineages: L4.1.2.1Haarlem, L4.1.1.1 X, L4.3.3 LAM, L4.2.2, L4.4.1 and L4.8 were identified. SIT 124, SIT 3234 and SIT 1952 belonging to undesignated lineage 4 and a new type were identified as L4.1.2.1 Haarlem and by phylogenetic analysis showed H3/SIT 49 was evolutionary closely linked to them. Novel 12 regions of difference (RDs) were detected and named them as SL-RDs. The absence of SL-RD 3,6,9 is a possible marker to identify locally circulating Haarlem strains in Sri Lanka. The clonal expansion of SIT 3234 was also confirmed by the phylogenetic analysis and differentiated into 2 clades based on SL-RD11. Deletion of SL-RD 11 in SIT 3234/clade II may have occurred as a local adaptation while evolution before the clonal expansion. SL-RD 11 could be a possible candidate for a specific genetic marker to differentially identify 2 clades of SIT 3234 together with other genotyping methods for continue monitoring to prevent future outbreaks. I found 123 non-synonymous SNPs in coding regions which were common to SIT 3234. Further analysis is required to identify the virulence properties and mechanisms of SIT 3234.

The population structure of MTB in Kandy, Sri Lanka was different from the South Asian Region. Clonal expansion of locally evolved lineage 4/SIT 3234 and detection of the pre-MDR Beijing isolates from new TB patients is alarming and will require continuous monitoring.