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

（スリランカにおいて肺結核患者より分離された結核菌株の分子疫学的解析）

Balapuwaduge Charitha Gayathri Mendis

TABLE OF CONTENT

ABBREVIATIONS

DR	Direct repeat
IS 6110 RFLP	Insertion sequence 61110 Restriction fragment length polymorphism
LSP	Large sequence polymorphism
MDR-TB	Multidrug resistant tuberculosis
MIRU-VNTR	Mycobacterial interspersed repetitive unit- variable number tandem repeats
MTB	*Mycobacterium tuberculosis*
MTBC	*Mycobacterium tuberculosis* complex
PCR	Polymerase chain reaction
QRDR	Quinolone resistance determining region
RD	Region of differences
SIT	Spoligo international type
SNP	Single nucleotide polymorphism
TB	Tuberculosis
TbD	*Mycobacterium tuberculosis* specific deletion
WHO	World Health Organization

PREFACE

Tuberculosis (TB) is one of the world leading cause of death by infectious disease with estimated 10 million incidences in 2017. It is mainly caused by *Mycobacterium tuberculosis* (MTB). Although TB is curable and preventable disease, due to incomplete understanding of the genetic variations of MTB that contribute to pathogenesis and antibiotic resistance, still we have not succeeded in combating MTB. It is believed that emergence of multidrug resistant TB (MDR-TB), HIV and poor TB control have contributed to the dramatic increase in the TB burden worldwide.

Sri Lanka is a moderate TB prevalence country in South Indian region. The TB incidence and mortality rates in Sri Lanka in 2017 were 64 and 3.2 per 100,000 population (World Health Organization, 2018). Sri Lanka also has a relatively good TB control programme, with a 69% case detection rate and 82.9% treatment success rate (World Health Organization, 2018). However, it is high time to think of more effective strategies to prevent and control TB in Sri Lanka to go in par with World Health Organization (WHO) End TB Strategy and end the TB epidemic by reducing TB deaths and new cases. Specially we have to target the interruption of transmission and prevention of the emergence of drug resistance outbreaks. One of the key factors that we need achieve this goal is national epidemiological data on circulating genotypes of MTB, transmission patterns, gene mutations conferring drug resistance in Sri Lanka.

MTB is one of the members of the *M. tuberculosis* complex (MTBC). It is an acid fast, intracellular, aerobic bacillus with a tough cell wall structure containing high content of mycolic acids, long chain cross linked fatty acids and other cell wall lipids. It is an intracellular pathogen and able to survive by slow growing in an adverse environment such as inside of the macrophage. MTB has a circular chromosome which contains about 4,200,000 base pairs consisiting of 65% GC content.

Genetic characterization of MTB has diversified the human-adapted strains into seven major lineages, which differ in their geographic distribution and association with human sub-populations (Gagneux et al. 2006b). They are lineage 1(Indo-Oceanic), lineage 2 (East-Asian), lineage 3 (East African-Indan), lineage 4 (Euro American), lineage 5 (West African 1), lineage 6 (West African 2) and lineage 7 ('Aethiops vetus'). Though MTB shows a strong phylogeographical population structure, some lineages occur globally while others show a strong geographical restriction. Therefore, understanding the genetic diversity of MTB strains in a given clinical setting is a key factor to inform the introduction of more effective control measures and patient management strategies.

Over the last decades different genotyping tools such as large sequence polymorphism (LSP), spoligotyping, mycobacterial interspersed repetitive unit- variable number tandem repeat (MIRU-VNTR) typing and whole genome sequencing have become beneficial in epidemiological studies by providing a platform to study the genetic diversity, transmission dynamics and phylogenetic analysis of MTB. LSP analysis is a PCR-based method that uses specific primers for the expected Regions of Difference (RD) for each lineage (Gagneux et al. 2006b). By performing LSP analysis MTB isolates can be assigned into lineage 1-6. Spoligotyping is a frequently used PCR-based molecular typing technique which allows the differentiation of MTB strains into different sub lineages. It uses a reverse-hybridization technique to detect variability in the direct repeat (DR) region which consists of multiple copies of a conserved 36-bp sequence separated by multiple unique spacer sequences in the genome of MTB (Kamerbeek et al. 1997). MIRU-VNTR uses the variability in the numbers of repeats present at particular tandem repeat loci in bacterial genomes, and involves PCR amplification of such tandem repeat loci and size calculation to identify the number of repeats at each locus in a given MTB strain (Supply et al. 2006). MIRU-VNTR method has been used along with spoligotyping as the combination of both approaches has more

discriminatory power to identify epidemiologically linked strains. With recent advances in next generation sequencing, the analysis of bacterial whole genome sequences has contributed significantly to the understanding of virulence factors and antibiotic resistance of MTB.

However, the exploration of molecular epidemiology of MTB in Sri Lanka is limited to several studies that have been performed using molecular DNA fingerprinting techniques such as IS*6110*-RFLP, spoligotyping, MIRU-VNTR (Rajapaksa et al. 2008; Magana-Arachchi et al. 2010, 2011; Weerasekera et al. 2015, 2019) and whole genome sequencing (Stucki et al. 2016). Therefore, we aimed to perform molecular characterization of MTB isolates from pulmonary tuberculosis patients in Sri Lanka in order to identify the population structure, transmission patterns and lineage 4 specific characteristics among a selected district (Kandy) in Sri Lanka.

Kandy is one of the main cities in Sri Lanka as well as the capital of Kandy District and the Central Province. It had 1,378,803 population and 3rd highest number of TB patients (n=720) in country in 2013. It is a hotspot for foreign and local pilgrims, tourists and traders since ancient time. Historically it is important as the last Sri Lankan monarchy, where mainly native Sri Lankan population resided until Sri Lanka became a dominion of the British Empire in 1815. During Portuguese (1517- 1638) and Dutch (1602- 1796) colonial periods they mainly interact with Kandy for trade, but British lived in Kandy. South Indians who came with the queens of South Indian origin to the Kandyan Kingdom and the Tamil plantation workers who were brought subsequently from South India to Central province by the British in mid-19th century also started living in and around Kandy District. With this background we selected Kandy District as our study site.

The present thesis consists of two chapters. The first chapter contains genotyping of MTB isolates from Kandy by spoligotyping, LSP analysis and MIRU-VNTR typing in order

to identify circulating genotypes of MTB and their transmission patterns within Kandy District , Sri Lanka. As we identified the predominant MTB lineage in Kandy District , Sri Lanka is lineage 4 (Euro-American lineage) and clonal expansion of locally evolved lineage 4/SIT3234 in chapter I, the focus of chapter II was lineage 4. In chapter II for a deeper understanding of the characteristics of lineage 4 specially concerning lineage 4/SIT3234, comparative genomic analysis was performed.

CHAPTER I

Insight into genetic diversity of *Mycobacterium tuberculosis* in Kandy Sri Lanka reveals predominance of the Euro- American lineage

Introduction

Tuberculosis (TB) is one of the ancient diseases known to mankind yet remains a major public health problem in many low- and middle-income countries. It has overtaken HIV/AIDS as the leading cause of death by a single infectious agent, with an estimated 10 million new TB cases with 1.6 million deaths worldwide in 2017. Two thirds of the estimated number of TB cases in 2017 occurred in Asian and African countries: India (27%), China (9%), Indonesia (8%), the Philippines (6%), Pakistan (5%), Nigeria (4%), Bangladesh (4%) and South Africa (3%). While India accounts for more than a quarter of the global TB burden, the neighboring country Sri Lanka (population 21 million) is among the moderate TB prevalence countries in the region. The TB incidence and mortality rates in Sri Lanka in 2017 were 64 and 3.2 per 100,000 population (World Health Organization, 2018).

It is believed that emergence of multidrug resistant TB (MDR-TB), HIV and poor TB control have contributed to the dramatic increase in the TB burden worldwide. In Sri Lanka, the estimated percentage of TB cases with MDR-TB among new TB patients was 0.5% while it was 4.1% among retreatment patients according to the national surveillance conducted in 2018. Sri Lanka also has a relatively good TB control programme, with a 69% case detection rate and 82.9% treatment success rate (World Health Organization, 2018). In addition, the low prevalence (less than 0.1%) of HIV/AIDS in Sri Lanka may also have contributed to it being an intermediate TB burden country.

However due to the changes in the sociocultural environment, an increasing prevalence of diabetes mellitus and use of immunosuppressive therapies, the TB situation in the country could change. Emigration and immigration could also change the current TB situation through the introduction of new *Mycobacterium tuberculosis* (MTB) strains which are more prone to develop drug resistance or more transmissible virulent. Hence monitoring the MTB population will provide important data to monitor and underpin the Sri Lankan TB control programme.

Genetic characterization of MTB has shown that the human-adapted strains are diversified into seven major lineages, which differ in their geographic distribution and association with human sub-populations (Gagneux et al. 2006b). Though MTB shows a strong phylogeographical population structure, some lineages occur globally while others show a strong geographical restriction. For example lineage 2 and 4 are widespread globally, probably due to high virulence, compared to lineage 5 and 6 which are highly restricted to West Africa; distinct lineages therefore appear to have differing propensities to transmit and develop drug resistance (Gagneux 2018). Therefore, understanding the genetic diversity of MTB strains in a given clinical setting is a key factor to inform the introduction of more effective control measures and patient management strategies.

Over the last decades different genotyping tools such as large sequence polymorphism (LSP), spoligotyping and mycobacterial interspersed repetitive unit- variable number tandem repeat (MIRU-VNTR) typing have become beneficial in epidemiological studies by providing a platform to study the genetic diversity, transmission dynamics and phylogenetic analysis of MTB.

LSP analysis is a PCR-based method that uses specific primers for the expected Regions of Difference (RD) for each lineage (Gagneux et al. 2006b). By performing LSP analysis MTB isolates can be assigned into lineage 1-6.

Spoligotyping is one of the most frequently used PCR based molecular typing techniques which allows the differentiation of MTB strains into different sub lineages/ clades. It is a hybridization assay that detects variability in the direct repeat (DR) region of the DNA of MTB (Kamerbeek et al. 1997). The DR region consists of multiple copies of a conserved 36-bp sequence (the DRs) separated by multiple unique spacer sequences. The entire DR locus is amplified by PCR using primers that are complementary to the sequence of the DRs. The PCR products are hybridized to a membrane with 43 spacer oligonucleotides. Each of the spacers produces either a dark band/spot (indicating the presence of the spacer) or no band/spot (indicating the spacer's absence). For each M. tuberculosis isolate, the spoligotyping assay produces a series of bands. The pattern is converted to a 43-digit binary code system that has 1s and 0s (1 means that the band is present and 0 means it is absent). The results can easily be interpreted and compared using SITVIT2 database. The sensitivity of spoligotyping is estimated to be 10 fg of chromosomal DNA.

MIRU-VNTR typing uses the variability in the numbers of repeats present at particular known tandem repeat loci in bacterial genomes. PCR amplification using primers specific for the regions of tandem repeat loci and the determination of the sizes of the amplicons, after electrophoretic migration are the two steps in this method. As the length of the repeat units is known, the sizes of the amplicons reflect the number of repeats in each locus. The final result is a numerical code, corresponding to the number of tandem repeats present in each locus and this serves as a DNA fingerprint of a bacterial isolate (Supply et al. 2006). MIRU-VNTR method has been used along with spoligotyping as the combination of both approaches has more discriminatory power to identify epidemiologically linked strains.

The molecular epidemiology of MTB is poorly explored in Sri Lanka. Although several studies have been performed that applied molecular DNA fingerprinting techniques such as IS*6110*-RFLP, spoligotyping and MIRU-VNTR (Rajapaksa et al. 2008; Magana-Arachchi et

al. 2011; Weerasekera et al. 2015, 2019), the results of these studies indicated the requirement for additional molecular epidemiological analysis of circulating genotypes of MTB in Sri Lanka. Therefore, this study was performed to identify circulating genotypes of MTB and their transmission patterns within Kandy district, in the Central Province in Sri Lanka by using spoligotyping, LSP analysis and MIRU-VNTR typing.

Materials and methods

Sample collection

The sputum samples were collected from 100 new pulmonary TB patients (patients with no evidence of past TB) who visited the Central Chest Clinic in Kandy Sri Lanka from December 2012 to October 2013. Only patients above 18 years of age and currently residing in Kandy district were included in this study. The collected sputum samples were processed and cultured on Lowenstein-Jensen medium at the Department of Microbiology in the Faculty of Medicine, University of Peradeniya. Data on patient demographics, risk factors and laboratory investigations were also collected. This study was ethically approved by the Ethical Review Committee, Faculty of Medicine, University of Peradeniya, Sri Lanka

DNA extraction

Suspected MTB colonies grown on Lowenstein-Jensen medium were suspended in 200µl of distilled water and heated for 20 minutes at 95°C. The heat killed bacteria were transported to Hokkaido University Research Center for Zoonosis Control in Japan and stored at -30°C. After several steps of freezing and boiling, the suspensions were centrifuged for 5 min at 10,000 rpm. Finally, the supernatant containing the bacterial DNA was retrieved and used for further molecular analysis.

Sequencing of drug resistance associating genes

Comparative sequence analysis of *rpoB* gene was performed to confirm the bacterial species (Helb et al. 2010; Poudel et al. 2012). Sequencing to detect mutations in genes associated with drug resistance was performed as described previously by Poudel et al. (2012), targeting the rifampicin resistance-determining region (RRDR) in *rpoB*, *katG* coding and *inhA*

regulatory regions, and the quinolone resistance-determining region (QRDR) in *gyrA* in order to identify multidrug resistant (MDR) and pre-extensively drug-resistant (pre-XDR) isolates. The sequences were compared with the wild-type sequences of H37Rv using BioEdit software version 7.0.9 (Hall, 1999). Phenotypic drug susceptibility test results were not available for these isolates.

Spoligotyping

All MTB isolates were analyzed by spoligotyping as previously described (Kamerbeek et al. 1997). The DR region in the mycobacterial genome was amplified by PCR, and the resulting products were hybridized to a set of 43 spacer specific oligonucleotide probes covalently bound to a membrane. Presence or absence of such spacer was determined (with a dark band indicating the presence of a spacer, while no band indicates a spacer's absence) and this pattern is converted to a 43-digit binary code system which was interpreted and compared using the SITVIT2 database (http://www.pasteur-guadeloupe.fr:8081/SITVIT2/) to determine the spoligotype international type (SIT) (Couvin et al. 2019).

Large sequence polymorphism (LSP)

MTB isolates of spoligotype patterns with either no assigned SIT or sub lineage were analyzed by LSP to assign lineages. PCR was performed using specific primers for the expected Regions of Difference, namely lineage 1-RD239, lineage 3- RD750, allowing lineages to be identified based on the size of PCR products as described by Gagneux et al. (2006b) and Tsolaki et al. (2004). Lineage 4 was identified based on the 7-bp deletion in *pks15/1* locus (Marmiesse et al. 2004).

MIRU-VNTR typing

MIRU-VNTR typing was performed by amplifying 24 loci, including 12 MIRU loci (MIRU2, MIRU4, MIRU10, MIRU16, MIRU20, MIRU23, MIRU24, MIRU26, MIRU27, MIRU31, MIRU39, and MIRU40), four exact tandem repeat (ETR) loci (ETR-A, ETR-B, ETR-C, and ETR-F), four Queens University Belfast (QUB) loci (QUB11a, QUB11b, QUB26, and QUB4156), and four VNTR loci (VNTR424, VNTR1955, VNTR2401, and VNTR3690) with modifications as described by Supply et al. (2006) for the selected clusters based on spoligotyping results. The number of tandem repeats for each locus was calculated from the PCR product size by conventional gel electrophoresis. Isolates that didn't show any band or showed multiple bands in more than two loci, suggestive of mixed infection, were excluded from the analysis after confirmation by repeat testing.

Data analysis

Statistical analysis was performed using RStudio (Integrated Development for R. RStudio, Inc., Boston, MA: URL http://www.rstudio.com/). Spoligoforest tree (Fruchterman-Reingold algorithm) was drawn using the spolTools online software (Reyes et al. 2008; Tang et al. 2008) available on http://spoltools.emi.unsw.edu.au/ to identify the evolutionary relationship among spoligotype patterns. A minimum spanning tree (MST) was constructed based on MIRU-VNTR results using BioNumerics software version 6.6 (Applied Maths, Belgium). Clusters were defined as two or more isolates sharing an identical 24-loci MIRU-VNTR pattern and the clustering rate was calculated using the formula: number of clustered isolates/total number of isolates (Glynn et al., 1999).

.

Results

MTB isolates

Out of 100 clinical isolates, 89 were confirmed as MTB by *rpoB* gene sequencing. As four isolates showed evidence of mixed infection with MTB in lineage 1 and 4, those were excluded. Finally, 85 isolates were used for molecular analysis. All the TB suspected patients living in the district were supposed to visit the Central Chest Clinic, thus, the 89 samples can be taken as representative of the region and were approximately 1/7 of the expected total TB incidence cases in the Kandy district during the collection period.

Drug resistance conferring gene mutations

Three isolates (3.5%), out of 85 were genotypically resistant to isoniazid. Two isolates had the G944C mutation (i.e. Ser315Thr substitution) in *katG* and one isolate had a mutation T-8A in the *inhA* regulatory region. No mutations were detected in the RRDR in *rpoB* and QRDR in *gyrA*.

Spoligotyping and LSP

Spoligotyping of 85 isolates enabled the detection of 26 distinct spoligotype patterns corresponding to 21 different SITs and 5 new patterns which have not been reported in SITVIT 2 database yet (Table 1). Those new patterns (New Type 1-5) were assigned into lineage 1 and 4 by LSP. The dominant lineage in our study was lineage 4 (n=39, 45.9%), followed by lineage 1 (n=25, 29.4%) and lineage 2 (n=20, 23.5%). We found only one isolate from lineage 3 (1.2%). The ratio of the lineage 4 was significantly higher than other lineages ($p<0.05$, Chi square test or Fisher's exact test). SIT1 (Beijing, lineage 2) was the most prevalent SIT found (n=19, 22.4%) followed by SIT11 (EAI3_IND, lineage 1; n=16, 18.8%), SIT124 (Undesignated, lineage 4; n=8, 9.4%) and SIT3234 (Undesignated, lineage 4; n=8, 9.4%) (Table 1). Two isolates from Beijing/SIT1 had a *katG* G944C mutation (Ser315Thr)

and one isolate from EAI 3_IND/ SIT355 had an *inhA* T-8A mutation.

MIRU-VNTR typing

Based on the spoligotyping results, clusters of: Beijing/SIT 1 (n=17/19; two isolates were excluded when constructing MST due to no bands in several loci); EAI3_IND/SIT 11 (n=16); Undesignated lineage 4/SIT 124 (n=8); and Undesignated lineage 4/SIT 3234 (n=8) were analyzed by 24 loci MIRU-VNTR typing and an MST was constructed (Fig. 2). The clustering rates in SIT 1, SIT 11 and SIT 124 were 41, 56 and 50%, respectively. All 8 isolates in SIT 3234 were in one cluster (clustering rate = 100%) together with 2 isolates of SIT 124. Genetically INH resistant isolates in SIT 1 (n=2) were singletons.

Analysis of patients' demographics, risk factors and laboratory findings

Complete data on patients' demographics, risk factors and laboratory findings (smear positivity and time to culture positivity) were available for 55 patients (Table 2). Overall, 42 patients were male and 13 were female (male to female ratio 3.23). The age of the patients ranged from 21-80 years. There was no significant association between variables and the lineage 4 or non-lineage 4 when compared category-wise.

Discussion

MTB, the main causative agent of human TB co-evolved with humans and its diversity has been shaped by human migration out of Africa and distinct human populations (Comas et al. 2013). By adapting to different human populations, lineage 1, 2, 3, 4 evolved and became endemic lineages in the Indian Ocean Region, East Asia, Central Asia and Europe respectively. Brosch et al (2002) found that MTB strains could be divided into “ancestral” and “modern” strains based on the presence or absence of an MTB specific deletion (TbD1) region. Among the four MTB lineages observed in our study, only the lineage 1 (labeled EAI or MANU in the spoligotyping nomenclature) possesses an intact TbD1 locus and is therefore an “ancient” type. Lineage 1 is suggested to have been the first MTB lineage that emerged out of Africa and became the predominant lineage in countries bordering the Indian ocean from Eastern Africa to Melanesia. Later, lineage 3 is thought to have emerged across Southern Asia and dispersed out of the Indian subcontinent (O’Neill et al. 2019). When the distribution of lineages was compared among different geographical areas of India, lineage 1 (EAI/TbD1+) was predominant in southern India while lineage 3 (CAS/TbD1-) was dominant in northern India. This suggests that lineage 1 could be the endemic lineage in Southern Asia, while lineage 3 emerged and spread from the northern to southern area in subsequent periods (Gutierrez et al. 2006; Thomas et al. 2011; Joseph et al. 2013; Varma-Basil et al. 2016; Manson et al. 2017a; Sharma et al. 2017) (Table 3, Table 4). In a previous study in Sri Lanka in which isolates were collected in Colombo, commercial capital on the west coast, lineage 1 was also reported as dominant with 58.2% of isolates belonging to this lineage (Rajapaksa et al. 2008). These findings are similar to those in the nearby region of southern India, suggesting that lineage 1 could be the endemic ‘domestic’ lineage in this location. Furthermore, the prevalence of lineage 3 was found to be less than 1% in Sri Lanka (Table 3, Table 4) suggesting less interaction between Sri Lanka and central or

northern India.

In contrast, our study revealed the predominant lineage circulating in the Kandy district was lineage 4, and not lineage 1 as expected. The historical relationship that Sri Lanka has had with European countries may have contributed to this finding. Sri Lanka was colonized by the Portuguese, Dutch and British for hundreds of years (16-17th, 17-18th and 19th-mid 20th century, respectively). We hence hypothesize that the introduction of lineage 4 into Sri Lanka may have happened during the European colonial period. Supporting our hypothesis, population genomic and phylogeographic analysis of MTB lineage 4 have found that dispersal of lineage 4 has been dominated by historical migrations out of Europe (Brynildsrud et al. 2018). This latter study demonstrated an intimate temporal relationship between European colonial expansion into Africa and the Americas and the spread of MTB lineage 4. In Sri Lanka, Portuguese and Dutch settlers mainly colonized the coastal area including Colombo, whereas British settlers scattered over the country and mainly resided in Kandy. Evidence for predominance of lineage 4 in the Kandy district may suggest that it was introduced as a founder MTB population, or alternatively that, as the "modern" lineage 4 (TbD1-) is suggested to have enhanced virulence and an ability to infect distinct human populations with different genetic backgrounds (Stucki et al. 2016), it may have outcompeted the "ancient" lineage 1 (TbD+) that may have been the endemic lineage in Kandy prior to colonization.

We identified 14 distinct spoligotype patterns in the lineage 4 isolates. Half of them were designated as Haarlem, T and X sublineages, which are well known to be prevalent in European countries. Comparison of these spoligopatterns with those circulating in other countries using data present in the SITVIT2 database revealed that SIT50, SIT49 and SIT53 have worldwide distribution including Portugal, Netherlands and the United Kingdom; SIT2 has been mainly distributed in Europe and SIT478 is prevalent in the European region. This

again provides circumstantial evidence that Portuguese, Dutch and British settlers introduced lineage 4 to Sri Lanka during the colonial period. SIT50 and SIT53 sublineages seem to be well established in Sri Lanka as they were also reported in previous studies (Rajapaksa et al. 2008; Weerasekera et al. 2015). The other half of the spoligotypes studied, in which the majority of lineage 4 isolates (26/39, 66.7%) were contained, were of a new or undesignated type. An important finding of this study was that 33.1% (27/85) of isolates had new or undesignated spoligotype patterns according to the SITVIT2 database and 96.3% (26/27) of those were identified as lineage 4 by LSP. This finding indicates that lineage 4 has been circulating in Kandy, Sri Lanka for a long time and that micro evolution to adapt to the Sri Lankan host population may have occurred. However, further detailed studies using techniques like such as whole genome sequencing and time-scaled haplotypic density are warranted to confirm the factors that shaped the local population structure of MTB in Sri Lanka.

A spoligoforest tree (Fig.2) revealed the probable parental links between the strains belonging to different sublineages. Most of ancestral lineage 1 (EAI) strains were linked within a parental network with no recent evolutionary connections to the new types. In contrast, the majority of lineage 4 strains were linked within a parental network together with undesignated and new types showing ongoing evolution. SIT124 is a probable descendent of SIT50 (Haarlem, H3) while SIT3234, SIT1952 and new type 1 have evolved from SIT124. MIRU-VNTR analysis using 24 loci showed all isolates (n=8) in SIT3234 were in one cluster together with 2 isolates of SIT124 indicating a clonal expansion of these sublineages in the study group. Out of eight SIT3234 isolates, four patients' demographic data were available and revealed that all the patients lived in different regions and that no direct contact between them was found suggesting this sub-lineage has already widely spread in the area. In the SITVIT2 database, 0.06% of isolates belong to SIT124 with a worldwide distribution that

includes India, China, Netherlands and United Kingdom, all of which are known to have deeply rooted historical relations with Sri Lanka. Previous studies also suggested the clonal expansion of this sublineage in Sri Lanka (Rajapaksa et al. 2008; Weerasekara et al. 2015). SIT3234 which was found in China (n=1) and France (n=1) in the SITVIT2 database was also reported in Sri Lanka (Weerasekara et al. 2015). Comparison of 15-loci MIRU-VNTR patterns of SIT124 and 3234 in our study with the SITVIT2 database revealed that identical or similar MIRU-VNTR patterns have not been reported previously. Therefore, clonal expansion of SIT3234 requires attention, monitoring and further characterization as it seems to have evolved in Sri Lanka with local adaptation. It also has a parental link with the Haarlem sublineage which is known to cause drug resistant epidemics (Mardassi et al. 2005; Khanipour et al. 2016; Tarashi et al. 2017). These SIT3234 isolates formed a cluster with Haarlem isolates in a NJ-Tree using 22 MIRU-VNTR loci in MIRU-VNTRplus (https://www.miru-vntrplus.org/MIRU/index.faces, Weniger et al. 2010) (data not shown). Evolutionary 'modern' sublineages like Beijing and Haarlem are suggested to be more virulent compared to 'ancient' ones such as EAI. Based on this assertion SIT124 and 3234, which showed clonal expansion in this study, could have implications for epidemiology and control of TB in Sri Lanka in the future.

The Beijing sublineage is considered to be one of the predominant MTB sublineages, with a worldwide distribution and particularly dominating in East and South East Asian countries (Tamaru et al. 2012; Merker et al. 2015). The Beijing sublineage is suggested to be more virulent than other sublineages, showing higher pathogenicity and increased mortality in animal studies (Parwati et al. 2010). This lineage also has a higher mutation rate which contributes to its success as a major sublineage responsible for MDR and XDR (Parwati et al. 2010; Merker et al. 2015; San et al. 2018). Ongoing transmission of the Beijing sublineage has previously been detected in Sri Lanka (Rajapaksa et al. 2008; Weerasekera et al. 2015),

as well as in our current study. While SIT1 is the most prevalent SIT that we have found, MIRU-VNTR results (Fig.1) showed highly diverse patterns. The Beijing lineage may have been introduced to Sri Lanka through trading links with South East Asian countries during the period that Sri Lanka was one of the main ports in ancient maritime silk and spice trade routes. Furthermore, continuous migration and emigration between populations in Sri Lanka, China and other South Asian countries that continues up to the present date may be responsible for the higher genetic diversity within this sublineage in Sri Lanka. In addition, there is a hypothesis that Beijing lineage strains may have spread as a result of their increased resistance to BCG induced immunity (Bifani et al. 2002), a suggestion which may also need to be considered for selective transmission of Beijing strains in Sri Lanka as it has a high coverage of BCG vaccination.

Two isolates from the Beijing/SIT1 clade had a G944C mutation (Ser315Thr) in *katG* suggesting resistant to isoniazid. The *katG* Ser315Thr mutation is a well-known low fitness cost substitution (Gagneux et al. 2006a; Manson et al. 2017b) that supports the maintenance of efficient transmission of drug resistant MTB and is associated with MDR epidemics worldwide (Manson et al. 2017b; Shah et al. 2017; San et al. 2018). The *katG* Ser315Thr mutation is reported to have arisen before mutations that conferred rifampicin resistance across all of the MTB lineages, geographical regions and time periods (Manson et al. 2017b). Monitoring the drug resistant patterns in TB patients in Sri Lanka is highly warranted so as to identify the trends in drug resistance, to inform current control and to prevent future outbreaks. Detection of the harbinger the mutation, katG Ser315Thr, also known as pre-MDR TB mutations, could be advantageous in this aspect.

Considering lineage 1, a high percentage of EAI3_IND/SIT11 was also observed in previous studies in Sri Lanka (Rajapaksa et al. 2008) and South India (Joseph et al. 2013), suggested that South India may represent the probable origin of this sublineage in Sri Lanka due to

migratory patterns that stretch back to ancient times. What's more, in a previous study by Rajapaksa et al (2008), the EAI5 sublineage was shown to be prevalent (n=20/98, 20%) in the Western province while present at a much lower prevalence in the Kandy district (n=3/85, 3.5%). These findings suggest that the diversity of the MTB population structure in Sri Lanka. A high proportion of the MANU sublineage was detected in Kandy by Weerasekera et al. (2015), but we were unable to identify any isolate within this sublineage. This discrepancy may occur because the MANU sublineage spoligotype could be constructed by combining more than two spoligotype patterns (Lazzarini et al. 2012; Diab et al. 2016) in situations of mixed infections or a contamination.

In summary, the predominant lineage of MTB in Kandy, Sri Lanka was lineage 4 which may have been introduced by European traders and settlers during the colonial period. As the isolates from lineage 4 were genetically diverse, with most of them were having an undesignated or new spoligotype pattern, we suggest that this lineage has circulated in Sri Lanka for a long time period with microevolution driving the emergence of new descendants which may be adapted to the local Sri Lankan host population. Therefore, the clonal expansion of locally evolved and potentially host-adapted undesignated lineage 4/SIT3234 requires continued monitoring to inform the control of current and future outbreaks. The Beijing/SIT1 clade was the most prevalent SIT found in this study indicating ongoing transmission that reflects the global situation with the Beijing lineage. Though we didn't find MDR TB in our study, two isolates of Beijing /SIT1 from new TB patients had well known pre-MDR *katG* G944C mutation (Ser315Thr) which alarms for monitoring. This study shows that it will be necessary to carry out continuous surveillance of genetic diversity and drug resistant TB to develop a clear picture of prevalence, transmission and evolution of the TB to prevent future epidemics in Sri Lanka.

Summary

Sri Lanka is a country where the molecular epidemiology of *Mycobacterium tuberculosis* (MTB) is poorly explored. Therefore, this study was performed to identify circulating lineages/sub lineages of MTB and their transmission patterns. DNA was extracted from 89 isolates of MTB collected during 2012 and 2013 from new pulmonary tuberculosis patients in Kandy, 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, 45.9%), followed by lineage 1 (Indo-Oceanic, 29.4%), lineage 2 (East-Asian, 23.5%) 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. 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.

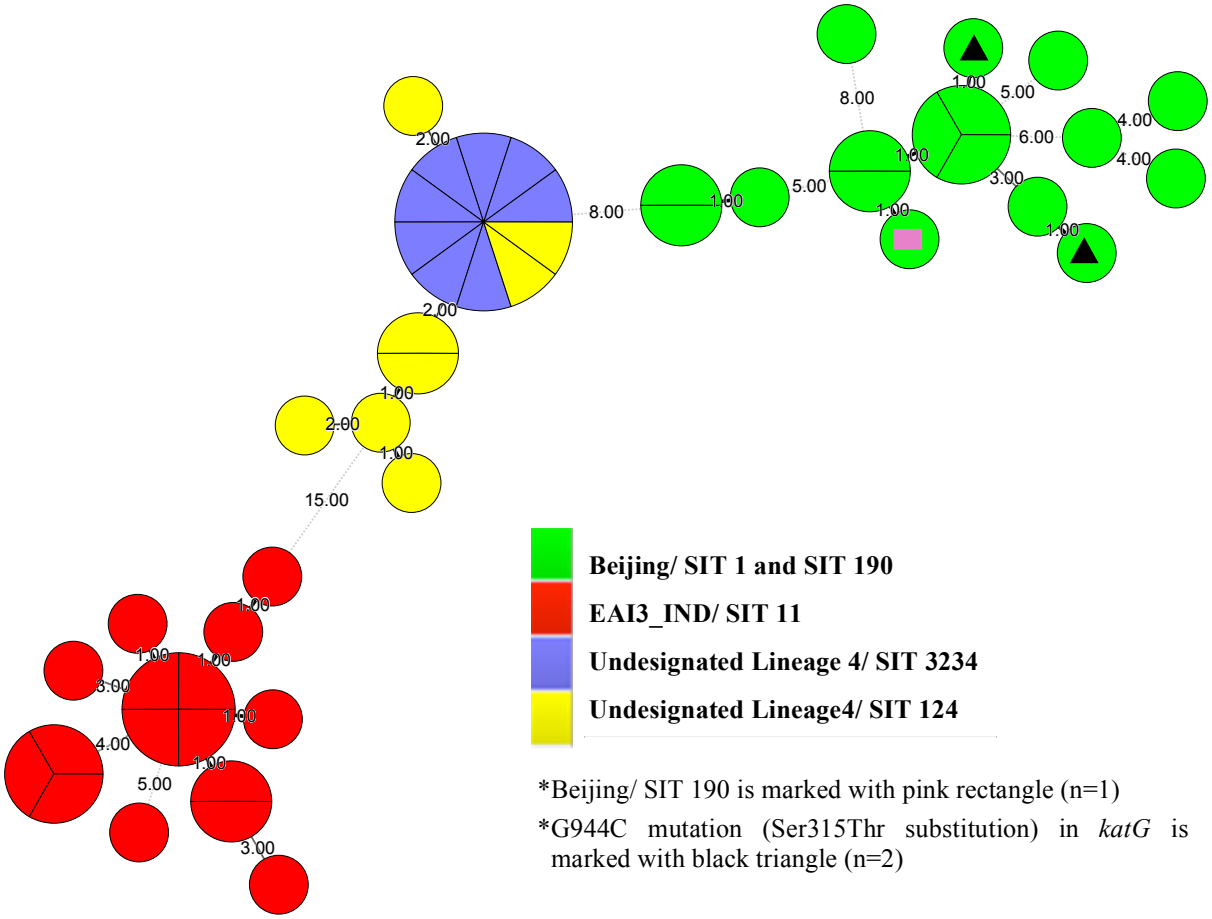


Figure 1. 24-loci MIRU-VNTR based MST of Beijing/ SIT1, EAI3_IND/ SIT11, Undesignated Lineage 4/ SIT124, Undesignated Lineage 4/ SIT 3234 isolates.

Each node represents a MIRU-VNTR type. The size of the node indicates the number of the isolates in each cluster. The length of the branches represents the distance between patterns while the numbers on the branch denotes the number of loci changes between two patterns. Green: Beijing/ SIT1, Red: EAI3_IND/ SIT11, Yellow: Undesignated Lineage 4/ SIT124, Purple: Undesignated Lineage 4/ SIT 3234

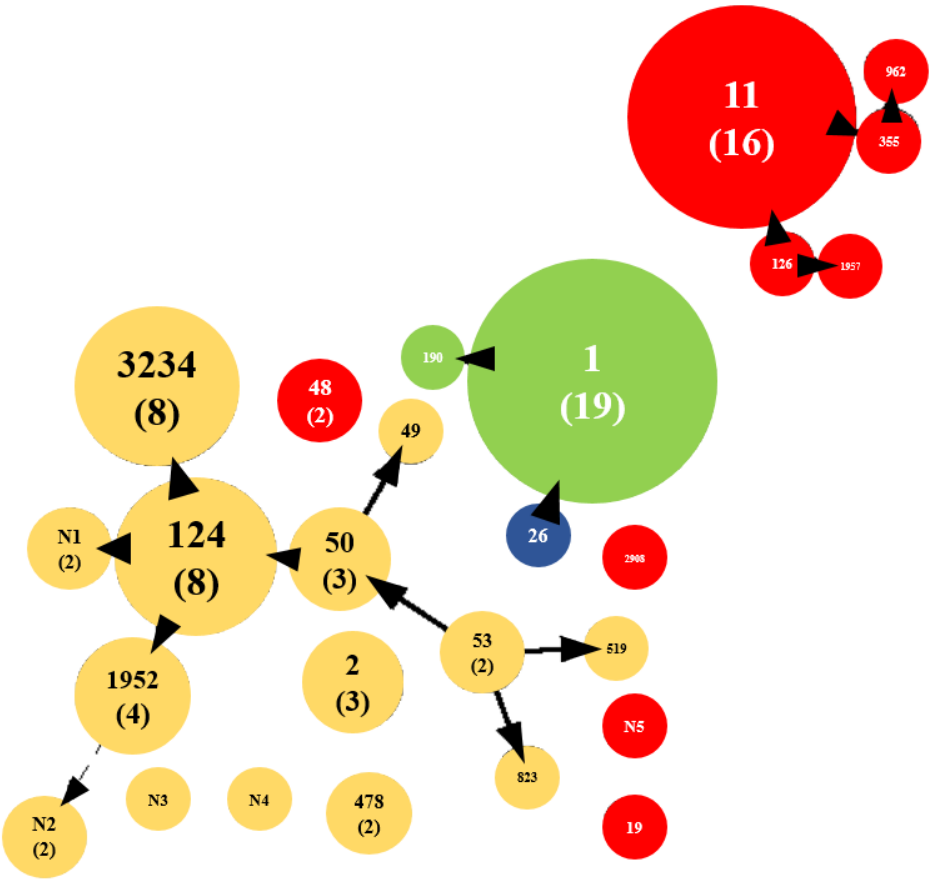
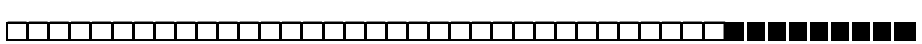
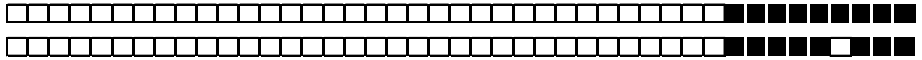
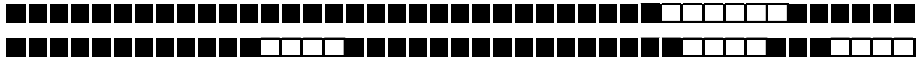
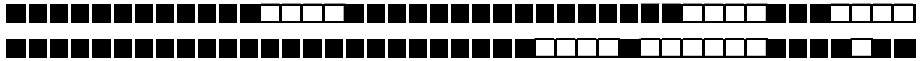


Figure 2. A Spoligoforest tree based on all spoligotypes

Each spoligotype pattern from the study is represented by a node with area size being proportional to the total number of isolates with that specific pattern. Changes (loss of spacers) are represented by directed edges between nodes, with the arrowheads pointing to descendant spoligotypes. The heuristic used selects a single inbound edge with a maximum weight using a Zipf model. Solid black lines link patterns that are very similar, i.e., loss of one spacer only (maximum weigh being 1.0), while dashed lines represent links of weight comprised between 0.5 and 1, and dotted lines a weight less than 0.5. Number inside the circle is SIT number while the number in parenthesis indicates the number of isolates in our study with that SIT.

Table 1

Description of 26 Spoligotype International Type (SITs; n=85 isolates) and corresponding spoligotyping defined sub lineages

Sub Lineage[a]	SIT [b]	Spoligotype Description[c]	Octal Number	No. of isolates	% of isolates
Lineage 1 (Indo-Oceanic Lineage)					
EAI3_IND	11		477777777413071	16	18.8
EAI1_SOM	48		777777777413731	2	1.2
EAI3_IND	355		477777777413031	1	1.2
EAI5	126		477777777413771	1	1.2
EAI5	962		777777777413031	1	1.2
EAI5	1957		477777777013771	1	1.2
EAI2_MANILLA	19		677777477413771	1	1.2
EAI6_BGD1	2908		777777757413671	1	1.2
New type 5			777775747413671	1	1.2
Lineage 2 (East-Asian Lineage)					
Beijing	1		000000000003771	19	22.4
Beijing	190		000000000003731	1	1.2
Lineage 3 (Central-Asian Lineage)					
CAS1-Delhi	26		703777740003771	1	1.2
Lineage 4 (Euro-American Lineage)					
Undesignated	124		777777777700771	8	9.4
Undesignated	3234		777777777600371	8	9.4
Undesignated	1952		777777774000771	4	4.7
H2	2		000000004020771	3	3.5
H3	50		777777777720771	3	3.5
T1	53		777777777760771	2	2.4
New type 1			777777777700671	2	2.4
X2	478		617776777760601	2	2.4
New type 2	orphan		777777774000731	2	2.4
H3	49		777777777720731	1	1.2
T1	823		776000003760771	1	1.2
T1	519		777777777740371	1	1.2
New type 3			777703777760700	1	1.2
New type 4			777777774100751	1	1.2

[a]Sub lineages were annotated using the SITVITWEB database

[b]SIT (Spoligo International types) were assigned by SITVITWEB database

[c]Closed squares represent positive hybridization (presence of spacer) and open squares represent no hybridization (absence of spacer)

(Undesignated - Spoligotype pattern is available in SITVIT 2 database and has SIT number, but the sub lineage is not assigned; Orphan- A SIT number has not been assigned for the Spoligotype pattern by SITVIT 2 database but has been reported by previous studies; New type- Spoligotype pattern found only in our study)

Table 2

Characteristics of patients (n=55) infected with Lineage 4 and Non-Lineage 4 isolates

	Variants	Number of Lineage 4 (n=30)	Number of non-Lineage 4 (n=25)	*p*-Value[a]
Demographics				
Gender	Male	26	16	0.0620
	Female	4	9	
Age	21- 40 years	6	11	0.1654
	41- 60 years	15	10	
	61-80 years	9	4	
Occupation[b]	Office workers/ businessman	8	6	
	Laborers	6	3	
	Estate workers	2	3	
	Drivers	1	2	
	Factory workers	0	2	
	Worked at Elderly home	0	1	
	Worked abroad	0	1	
	Worked in Tourist industry	1	2	
	Housewife	2	5	
	Non respondent	10	0	
Risk factors				
Smoking	Yes	15	12	1.0000
	No	15	13	
Comorbidity	Diabetes mellitus	10	7	0.9268
	Other comorbidities (lung disease, taking chemotherapy, cancer)	4	3	
	None	16	15	
Contact history	Yes	2	1	1.0000
	No	28	24	
Laboratory				

findings				
Direct smear	Positive	19	18	0.5716
	Negative	11	7	
Time for become culture positive	1-2 weeks	8	4	0.3466
	3-4 weeks	16	11	
	5-6 weeks	3	7	
	7-8 weeks	3	3	

[a]A *p*-value of <0.05 was considered significant; determined by Fisher's exact test.

[b]As number of patients in each variable of "occupation" were low, it was excluded from statistical analysis

Table 3 Summary of *M.tuberculosis* lineages distribution from previous studies in Sri Lanka and India

		Country of isolation %							
		Sri Lanka		India					
		Current study 2019 N=100	Rajapaksa et al. (2008) N= 98	Joseph et al. (2013) N=168	Manson et al. (2017a) N=201	Thomas et al. (2011) N= 101	Sharma et al. (2017) N= 335	Varma-Basil et al. (2016) N= 139	Gutierrez et al. (2006) N=91
Period of sample collection		2012-2013	2005-2006	1998-2005	1999-2005	2000-2005	2005-2007	2010-2011	1996- 2002
Geographical area		Kandy District	Colombo[a]	Kerala[b]	Tamil Nadu[b]	Andhra Pradesh[c]	Ghatampur[d]	Delhi[d]	12 different region
Lineages	**Lineage 1**	29.6	58.2	81.5	70.0	48.5	22.4	23.0	45,0
	Lineage 2	23.6	14.3	2.4	11.0	4.0	3.9	6.5	10.0
	Lineage 3	1.2	0	6.5	16.0	40.6	63.6	53.2	26.0
	Lineage 4	46.1	27.6	9.5	3.0	6.9	10.1	14.4	19.0

[a]Sample collection site was Colombo, however, the residences of patients were unclear [b]Kerala and Tamil Nadu represent Southern India.
[c]Andhra Pradesh represent South Eastern India. [d]Ghatampur and Delhi represent Northern India. [e]U = Undesignated

Table 4 Summary of *M.tuberculosis* sublineages (selected) distributions from previous studies in Sri Lanka and India

		Country of isolation %							
		Sri Lanka		India					
		Current study (2019) N=100	Rajapaksa et al. (2008) N= 98	Joseph et al. (2013) N=168	Manson et al. (2017a) N=201	Thomas et al. (2011) N= 101	Sharma et al. (2017) N= 335	Varma-Basil et al. (2016) N= 139	Gutierrez et al. (2006) N=91
Period of sample collection		2012-2013	2005-2006	1998- 2005	1999- 2005	2000- 2005	2005- 2007	2010-2011	1996-2002
Geographical area		Kandy District	Colombo[a]	Kerala[b]	Tamil Nadu[b]	Andhra Pradesh[c]	Ghatampur[d]	Delhi[d]	12 different region
Lineages/ sublineages									
Lineage 1	**EAI**	28.4	33.6	64.9	70	20.8	19.1	14.4	32.9
	Manu	0	0	1.2	0	7.9	3.0	1.4	0
Lineage 2	**Beijing**	23.6	14.3	2.4	11	2.0	3.3	6.5	8.5
Lineage 3	**CAS**	1.2	0	4.8	16	26.7	59.1	48.2	19.2
Lineage 4	**Haarlem**	8.2	3.1	1.		2.8	0.6	2.8	0
	T	4.8	5.1	3.6	1.5	1.0	5.07	7.2	4.3
	X	2.4	0	0	0	0	2.68	1.4	0
	LAM	0	1.0	0.6	0	0	0.29	0	0
	U[e]/SIT 124	9.4	6.1	3.0	0	0	0	0	0
	U[e]/SIT 3234	9	0	0	0	0	0	0	0
	U[e]/SIT 1952	4.7	0	1.2	0	0	0.3	0	0

[a]Sample collection site was Colombo, however, the residences of patients were unclear [b]Kerala and Tamil Nadu represent Southern India.

[c]Andhra Pradesh represent South Eastern India. [d]Ghatampur and Delhi represent Northern India. [e]U = Undesignated

DISSERTATION SUMMARY

TB is a major public health problem worldwide with no exception in Sri Lanka. Although Sri Lanka is a moderate TB prevalent country in South Indian region, it is high time to think of more effective strategies to prevent and control TB in Sri Lanka to end the TB epidemic by reducing TB deaths and new cases. One of the key factors that we need achieve this goal is epidemiological data on circulating genotypes of MTB, transmission patterns, gene mutations conferring drug resistance. As the exploration of molecular epidemiology of MTB in Sri Lanka is limited to several studies, we aimed to perform molecular characterization of MTB isolates from pulmonary tuberculosis patients in Kandy district, Sri Lanka. First we identified MTB lineage 4 is playing a major role in TB burden in Kandy district in Chapter 1. And in chapter II, we performed whole genome sequencing selecting 20 isolates of lineage 4 to get deep understanding on genetic diversity. Based on sublineage specific deletion of RDs and SNPs, six sublineages were identified and the majority were L4.1.2.1Haarlem. Previously unreported 12 RDs were detected among lineage 4 isolates. Out of them combination of SL-RD 3,6,9 could be used as a marker to identify locally circulating Haarlem strains in Sri Lanka. The clonal expansion of SIT 3234 which was notice in chapter I, was confirmed by the phylogenetic analysis and identified two clades of SIT 3234 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. We 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.

When we combined the results of spoligotyping and whole genome sequencing the most common sublineage in Kandy was Haarlem (34.1%) followed by EAI (28.4%) and Beijing

(23.6%). Haarlem sublineage is well known to cause out breaks and drug resistant TB mainly in European countries.

I believe this study underlines the need for continuous surveillance of genetic diversity and drug resistant MTB so as to develop a clear picture of prevalence, transmission and evolution of MTB that can underpin the current TB control programme and prevent future epidemics in Sri Lanka.

ACKNOWLEDGEMENT

It is a great pleasure for me to express my sincere gratitude to my supervisor Prof. Yasuhiko Suzuki from Division of Bioresources, Hokkaido University Research Center for Zoonosis Control for providing me an excellent opportunity to joined his laboratory and for his guidance and continuous support throughout my PhD study and related research. I greatly appreciate for great support and intellectual advice of Assoc. Prof. Chie Nakajima from Division of Bioresources, Hokkaido University Research Center for Zoonosis Control. Without their encouragement and warm support, I won't be able to succeed in my PhD study. I'm also indebted for their outstanding mentorship and inspiration which encourage me to be a good teacher and researcher like them.

I owe a deep sense of gratitude to Prof. Hideaki Higashi and Prof. Norikazu Isoda for their invaluable guidance and suggestions to improve my study. Their inspiration and timely suggestions with kindness have enabled me to complete my thesis too.

My sincere gratitude goes to Prof. D.B.M. Wickramarathne, former Dean, Faculty of Allied Health Sciences, University of Peradeniya, Sri Lanka and Dr. M.P.S. Mudalige, Head, Department of Medical Laboratory Science, Faculty of Allied Health Sciences, University of Peradeniya, Sri Lanka for encouraging and allowing me to study abroad with greater consideration on the future prospect.

I would like to express my thanks to Dr. Chandika Gamage from Department of Microbiology, Faculty of Medicine, University of Peradeniya, Sri Lanka for kindly introducing and encouraging me to apply for PhD programme in Hokkaido University under the guidance of Prof. Yasuhiko Suzuki, an excellent supervisor. I also grateful to Dr. Champa Ranatunga for her invaluable contribution in my study.

I would like to thank all members of Division of Bioresources for their kind support throughout my study period. Specially Ms. Yukari Fukushima for technical support in all lab

work and Ms. Yuko Hidaka for continuous support and encouraging with a kind heart to move on.

My sincere thanks go to all my lab mates: Dr. Ruchirada Changkwanyeun, Dr. Marvin Ardeza Villaneuea, Dr. Tomoyuki Yamaguchi, Dr. Jeewan Thapa, dr. Yogendra Shah, Dr. Nan Aye Thida Oo, Dr. Lai Lai San, Mr. Jong- Hoon Park, Mr. Yuki Ouchi, Mr. Kentaro Koide, Ms. Dipti Shrestha, Ms. Ruttana Pachanon, Mr. Thoko Flav Kapalamula, Ms. Mwangala Lonah Akapelwa, Ms. Risa Tsunoda, Ms. Wimonrat Tanomsridachchai, Mr. Lawrence P. Belotindos, Ms. Precious Bwalya, Mr. Joseph Yamweka Chizimu, Ms. Nami Ajima and other friends in Hokkaido University Research Center for Zoonosis Control for their kind support and making a friendly working environment.

Last but not least I would like to thank specially my mother, husband and my little son whose many sacrifices led to my successful completion of this four-year journey to obtain PhD. I also thank to my sisters and other family members for their moral support to continue my study and for always trusting me.

REFERENCES

Anh DD, Borgdorff MW, Van LN, Lan NTN, Van Gorkom T, Kremer K, et al. Mycobacterium tuberculosis Beijing genotype emerging in vietnam. Emerg Infect Dis. 2000;6(3):302–5.

Bifani PJ, Mathema B, Kurepina NE, Kreiswirth BN. Global dissemination of the Mycobacterium tuberculosis W-Beijing family strains. Trends Microbiol. 2002;10(1):45–52.

Brudey K, Driscoll JR, Rigouts L, Prodinger WM, Gori A, Al-Hajoj SA, et al. Mycobacterium tuberculosis complex genetic diversity: Mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. BMC Microbiol. 2006;6:1–17.

Coll F, McNerney R, Guerra-Assunção JA, Glynn JR, Perdigão J, Viveiros M, et al. A robust SNP barcode for typing Mycobacterium tuberculosis complex strains. Nat Commun. 2014;5:4–8.

Comas I, Coscolla M, Luo T, Borrell S, Holt KE, Kato-Maeda M, et al. Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans. Nat Genet [Internet]. 2013;45(10):1176–82. Available from: http://dx.doi.org/10.1038/ng.2744

Comas I, Homolka S, Niemann S, Gagneux S. Genotyping of genetically monomorphic bacteria: DNA sequencing in Mycobacterium tuberculosis highlights the limitations of current methodologies. PLoS One. 2009;4(11).

Coscolla M, Gagneux S. Consequences of genomic diversity in mycobacterium tuberculosis. Semin Immunol [Internet]. 2014;26(6):431–44. Available from: http://dx.doi.org/10.1016/j.smim.2014.09.012

Couvin D, David A, Zozio T, Rastogi N. Macro-geographical specificities of the prevailing tuberculosis epidemic as seen through SITVIT2, an updated version of the Mycobacterium tuberculosis genotyping database. Infect Genet Evol [Internet]. 2019;(October):1–13. Available from: https://doi.org/10.1016/j.meegid.2018.12.030

Couvin D, Rastogi N. Tuberculosis - A global emergency: Tools and methods to monitor, understand, and control the epidemic with specific example of the Beijing lineage. Tuberculosis [Internet]. 2015;95(S1):S177–89. Available from: http://dx.doi.org/10.1016/j.tube.2015.02.023

Cubillos-Ruiz A, Sandoval A, Ritacco V, López B, Robledo J, Correa N, et al. Genomic signatures of the Haarlem lineage of Mycobacterium tuberculosis: Implications of strain genetic variation in drug and vaccine development. J Clin Microbiol. 2010;48(10):3614–23.

Cui T, Zhang L, Wang X, He ZG. Uncovering new signaling proteins and potential drug

targets through the interactome analysis of Mycobacterium tuberculosis. BMC Genomics. 2009;10:1–10.

Diab HM, Nakajima C, Kotb SA, Mokhtar A, Khder NFM, Abdelaal ASA, et al. First insight into the genetic population structure of Mycobacterium tuberculosis isolated from pulmonary tuberculosis patients in Egypt. Tuberculosis [Internet]. 2016;96:13–20. Available from: http://dx.doi.org/10.1016/j.tube.2015.11.002

Felsenstein J. Confidence Limits on Phylogenies: an Approach Using the Bootstrap. Evolution [Internet]. 1985;39(4):783–91. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28561359

Gagneux S. Ecology and evolution of Mycobacterium tuberculosis. Nat Rev Microbiol [Internet]. 2018;16(4):202–13. Available from: http://dx.doi.org/10.1038/nrmicro.2018.8

Gagneux S, Burgos M V., DeRiemer K, Enciso A, Muñoz S, Hopewell PC, et al. Impact of bacterial genetics on the transmission of isoniazid-resistant Mycobacterium tuberculosis. PLoS Pathog. 2006a;2(6):0603–10.

Gagneux S, DeRiemer K, Van T, Kato-Maeda M, de Jong BC, Narayanan S, et al. Variable host-pathogen compatibility in Mycobacterium tuberculosis. Proc Natl Acad Sci [Internet]. 2006b;103(8):2869–73. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.0511240103

Gutierrez MC, Ahmed N, Willery E, Narayanan S, Hasnain SE, Chauhan DS, et al. Predominance of Ancestral Lineages of *Mycobacterium tuberculosis* in India. Emerg Infect Dis [Internet]. 2006;12(7):1367–74. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17073085%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3294724%5Cnhttp://wwwnc.cdc.gov/eid/article/12/09/05-0017_article.htm

Joseph B V., Soman S, Radhakrishnan I, Hill V, Dhanasooraj D, Ajay Kumar R, et al. Molecular epidemiology of Mycobacterium tuberculosis isolates from Kerala, India using IS6110-RFLP, spoligotyping and MIRU-VNTRs. Infect Genet Evol [Internet]. 2013;16:157–64. Available from: http://dx.doi.org/10.1016/j.meegid.2013.01.012

Kamerbeek J, Schouls LEO, Kolk A, Kuijper S, Bunschoten A, Molhuizen H, et al. Simultaneous Detection and Strain Differentiation of. J Clin Microbiol. 1997;35(4):907–14.

Khanipour S, Ebrahimzadeh N, Masoumi M, Sakhaei F, Alinezhad F, Safarpour E, et al. Haarlem 3 is the predominant genotype family in multidrug-resistant and extensively drug-resistant Mycobacterium tuberculosis in the capital of Iran: A 5-year survey. J Glob Antimicrob Resist [Internet]. 2016;5:7–10. Available from: http://dx.doi.org/10.1016/j.jgar.2016.01.007

Kremer K, Frothingham R, Haas WH, Hermans PWM, Palittapongarnpim P, Plikaytis BB, et al. Comparison of Methods Based on Different Molecular Epidemiological Markers for

Typing of Mycobacterium tuberculosis Complex Strains. J Clin Microbiol. 1999;37(8):2607–18.

Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol [Internet]. 2016;33(7):1870–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27004904

Lazzarini LCO, Rosenfeld J, Huard RC, Hill V, Lapa e Silva JR, DeSalle R, et al. Mycobacterium tuberculosis spoligotypes that may derive from mixed strain infections are revealed by a novel computational approach. Infect Genet Evol [Internet]. 2012;12(4):798–806. Available from: http://dx.doi.org/10.1016/j.meegid.2011.08.028

Magana-arachchi D. Pattern of Circulating Mycobacterium tuberculosis Strains in Sri Lanka. Tuberc - Glob Exp Innov Approaches to Diagnosis. 1999;511–26.

Magana-Arachchi DN, Medagedara D, Thevanesam V. Molecular characterization of Mycobacterium tuberculosis isolates from Kandy, Sri Lanka. Asian Pacific J Trop Dis [Internet]. 2011;1(3):181–6. Available from: http://dx.doi.org/10.1016/S2222-1808(11)60024-8

Magana-Arachchi DN, Perera AJ, Senaratne V, Chandrasekharan N V. Patterns of drug resistance and RFLP analysis of Mycobacterium tuberculosis strains isolated from recurrent tuberculosis patients in Sri Lanka. Southeast Asian J Trop Med Public Health. 2010;41(3):583–9.

Manson AL, Abeel T, Galagan JE, Sundaramurthi JC, Salazar A, Gehrmann T, et al. Mycobacterium tuberculosis whole genome sequences from Southern India suggest novel resistance mechanisms and the need for region-specific diagnostics. Clin Infect Dis. 2017a;64(11):1494–501.

Manson AL, Cohen KA, Abeel T, Desjardins CA, Armstrong DT, Barry CE, et al. Genomic analysis of globally diverse Mycobacterium tuberculosis strains provides insights into the emergence and spread of multidrug resistance. Nat Genet [Internet]. 2017b;49(3):395–402. Available from: http://dx.doi.org/10.1038/ng.3767

Mardassi H, Namouchi A, Haltiti R, Zarrouk M, Mhenni B, Karboul A, et al. Tuberculosis due to resistant Haarlem strain, Tunisia. Emerg Infect Dis. 2005;11(6):957–61.

Marmiesse M, Brodin P, Buchrieser C, Gutierrez C, Simoes N, Vincent V, et al. Macro-array and bioinformatic analyses reveal mycobacterial “core” genes, variation in the ESAT-6 gene family and new phylogenetic markers for the Mycobacterium tuberculosis complex. Microbiology. 2004;150(2):483–96.

Merker M, Blin C, Mona S, Duforet-Frebourg N, Lecher S, Willery E, et al. Evolutionary history and global spread of the Mycobacterium tuberculosis Beijing lineage. Nat Genet [Internet]. 2015;47(3):242–9. Available from: http://dx.doi.org/10.1038/ng.3195

Mokrousov I. Insights into the origin, emergence, and current spread of a successful Russian clone of Mycobacterium tuberculosis. Clin Microbiol Rev. 2013;26(2):342–60.

Parwati I, van Crevel R, van Soolingen D. Possible underlying mechanisms for successful emergence of the Mycobacterium tuberculosis Beijing genotype strains. Lancet Infect Dis [Internet]. 2010;10(2):103–11. Available from: http://dx.doi.org/10.1016/S1473-3099(09)70330-5

Poudel A, Nakajima C, Fukushima Y, Suzuki H, Pandey BD, Maharjan B, et al. Molecular Characterization of Multidrug-Resistant Mycobacterium tuberculosis Isolated in Nepal. Antimicrob Agents Chemother. 2012;56(6):2831–6.

Qian L, Haas PEWDE, Douglas JT, Traore H, Portaels F, Qing HZI, et al. Predominance of a Single Genotype of Mycobacterium tuberculosis in Countries of East Asia. J Clin Microbiol. 2000;33(12):3234–8.

Rajapaksa US, Perera AJ. Sublineages of Beijing Strain of Mycobacterium tuberculosis in Sri Lanka. Indian J Microbiol. 2011;51(3):410–2.

Rajapaksa US, Victor TC, Perera AJ, Warren RM, Senevirathne SM. Molecular diversity of Mycobacterium tuberculosis isolates from patients with pulmonary tuberculosis in Sri Lanka [Internet]. Vol. 102, Trans R Soc Trop Med Hyg. 2008. p. 997–1002. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=18513770

Rao KR, Kauser F, Srinivas S, Zanetti S, Sechi LA, Ahmed N, et al. Analysis of Genomic Downsizing on the Basis of Region-of-Difference Polymorphism Profiling of Mycobacterium tuberculosis.pdf. 2005;43(12):5978–82.

Reyes JF, Francis AR, Tanaka MM. Models of deletion for visualizing bacterial variation: An application to tuberculosis spoligotypes. BMC Bioinformatics. 2008;9:1–16.

Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol [Internet]. 1987;4(4):406–25. Available from: http://www.ncbi.nlm.nih.gov/pubmed/3447015

San LL, Aye KS, Oo NAT, Shwe MM, Fukushima Y, Gordon S V, et al. Insight into multidrug-resistant Beijing genotype Mycobacterium tuberculosis isolates in Myanmar. Int J Infect Dis [Internet]. 2018;76:109–19. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1201971218344448

Shah Y, Maharjan B, Thapa J, Poudel A, Diab HM, Pandey BD, et al. High diversity of multidrug-resistant Mycobacterium tuberculosis Central Asian Strain isolates in Nepal. Int J Infect Dis [Internet]. 2017;63:13–20. Available from: http://dx.doi.org/10.1016/j.ijid.2017.06.010

Sharma P, Katoch K, Chandra S, Chauhan DS, Sharma VD, Couvin D, et al. Comparative study of genotypes of Mycobacterium tuberculosis from a Northern Indian setting with strains reported from other parts of India and neighboring countries. Tuberculosis [Internet]. 2017;105:60–72. Available from:

http://dx.doi.org/10.1016/j.tube.2017.04.003

Singh J, Sankar MM, Kumar P, Couvin D, Rastogi N, Singh S, et al. Genetic diversity and drug susceptibility profile of Mycobacterium tuberculosis isolated from different regions of India. J Infect [Internet]. 2015;71(2):207–19. Available from: http://dx.doi.org/10.1016/j.jinf.2015.04.028

Stucki D, Brites D, Jeljeli L, Coscolla M, Liu Q, Trauner A, et al. Mycobacterium tuberculosis lineage 4 comprises globally distributed and geographically restricted sublineages. Nat Genet. 2016;48(12):1535–43.

Supply P. Multilocus Variable Number Tandem Repeat Genotyping of Mycobacterium tuberculosis. Inst Pasteur Lille [Internet]. 2005;(May):73. Available from: http://www.miru-vntrplus.org/MIRU/miruinfo.faces

Supply P, Rastogi N, Kreiswirth B, Locht C, Kurepina N, van Deutekom H, et al. Proposal for Standardization of Optimized Mycobacterial Interspersed Repetitive Unit-Variable-Number Tandem Repeat Typing of Mycobacterium tuberculosis. J Clin Microbiol. 2006;44(12):4498–510.

Tamaru A, Nakajima C, Wada T, Wang Y, Inoue M, Kawahara R, et al. Dominant Incidence of Multidrug and Extensively Drug-Resistant Specific Mycobacterium tuberculosis Clones in Osaka Prefecture, Japan. PLoS One. 2012;7(8):3–9.

Tamura K, Kumar S. Evolutionary distance estimation under heterogeneous substitution pattern among lineages. Mol Biol Evol. 2002;19(10):1727–36.

Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol [Internet]. 1993;10(3):512–26. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8336541

Tanaka MM, Francis AR. Detecting emerging strains of tuberculosis by using spoligotypes. Proc Natl Acad Sci. 2006;103(41):15266–71.

Tang C, Reyes JF, Luciani F, Francis AR, Tanaka MM. spolTools: Online utilities for analyzing spoligotypes of the Mycobacterium tuberculosis complex. Bioinformatics. 2008;24(20):2414–5.

Tarashi S, Fateh A, Rahimi Jamnani F, Siadat SD, Vaziri F. Prevalence of Beijing and Haarlem genotypes among multidrug-resistant Mycobacterium tuberculosis in Iran: Systematic review and meta-analysis. Tuberculosis [Internet]. 2017;107(358):31–7. Available from: https://doi.org/10.1016/j.tube.2017.03.005

Thomas SK, Iravatham CC, Moni BH, Kumar A, Archana B V., Majid M, et al. Modern and ancestral genotypes of mycobacterium tuberculosis from Andhra Pradesh, India. PLoS One. 2011;6(11):2–6.

Tsolaki AG, Hirsh AE, DeRiemer K, Enciso JA, Wong MZ, Hannan M, et al. Functional and evolutionary genomics of Mycobacterium tuberculosis: Insights from genomic deletions in 100 strains. Proc Natl Acad Sci. 2004a;101(14):4865–70.

Tsolaki AG, Hirsh AE, DeRiemer K, Enciso JA, Wong MZ, Hannan M, et al. Functional and evolutionary genomics of Mycobacterium tuberculosis: Insights from genomic deletions in 100 strains. Proc Natl Acad Sci. 2004b;101(14).

Universit N, Edition NS. 1, 2 , 1 2 ,. 2009;34(2):51–61.

Varma-Basil M, Narang A, Chakravorty S, Garima K, Gupta S, Kumar Sharma N, et al. A snapshot of the predominant single nucleotide polymorphism cluster groups of Mycobacterium tuberculosis clinical isolates in Delhi, India. Tuberculosis [Internet]. 2016;100:72–81. Available from: http://dx.doi.org/10.1016/j.tube.2016.07.007

Wang J, Liu Y, Zhang CL, Ji BY, Zhang LZ, Shao YZ, et al. Genotypes and characteristics of clustering and drug susceptibility of mycobacterium tuberculosis isolates collected in heilongjiang province, China. J Clin Microbiol. 2011;49(4):1354–62.

Weerasekera D, Magana-arachchi D, Madegedara D, Dissanayake N, Thevanesam V. Asian Pacific Journal of Tropical Disease. Asian Pacific J Trop Dis [Internet]. 2015;5(5):385–92. Available from: http://dx.doi.org/10.1016/S2222-1808(14)60802-1

Weerasekera D, Pathirane H, Madegedara D, Dissanayake N, Thevanesam V, Magana-Arachchi DN. Evaluation of the 15 and 24-loci MIRU-VNTR genotyping tools with spoligotyping in the identification of Mycobacterium tuberculosis strains and their genetic diversity in molecular epidemiology studies. Infect Dis (Auckl) [Internet]. 2019;0(0):1–10. Available from: https://doi.org/10.1080/23744235.2018.1551619

Comparing Genomes within the Species Mycobacterium tuberculosis. Genome Res. 11(4):547–54.

World Health Organization. (2018). Global tuberculosis report 2018. World Health Organization. http://www.who.int/iris/handle/10665/274453. License: CC BY-NC-SA 3.0 IGO

1