International Journal of Infectious Diseases 87 (2019) 84–91

Contents lists available at ScienceDirect

International Journal of Infectious Diseases

journal homepage: www.elsevier.com/locate/ijid

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

Charitha Mendis[a,b], Vasanthi Thevanesam[c], Athula Kumara[c], Susiji Wickramasinghe[d], Dushantha Madegedara[e], Chandika Gamage[c], Stephen V. Gordon[f,g], Yasuhiko Suzuki[a,g], Champa Ratnatunga[c,**], Chie Nakajima[a,g,*]

[a] Division of Bioresources, Hokkaido University Research Center for Zoonosis Control, Sapporo, Japan
[b] Department of Medical Laboratory Science, Faculty of Allied Health Sciences, University of Peradeniya, Sri Lanka
[c] Department of Microbiology, Faculty of Medicine, University of Peradeniya, Sri Lanka
[d] Department of Parasitology, Faculty of Medicine, University of Peradeniya, Sri Lanka
[e] Respiratory Disease Treatment Unit, Teaching Hospital, Kandy, Sri Lanka
[f] School of Veterinary Medicine, University College Dublin, Belfield, Dublin, Ireland
[g] Global Station for Zoonosis Control, Hokkaido University Global Institute for Collaborative Research and Education, Sapporo, Japan

ARTICLE INFO

Article history:
Received 17 May 2019
Received in revised form 24 June 2019
Accepted 1 July 2019
Corresponding Editor: Eskild Petersen, Aarhus, Denmark

Keywords:
Mycobacterium tuberculosis
Spoligotype
MIRU-VNTR
Lineage 4 (Euro-American lineage)
Sri Lanka

ABSTRACT

Objective: 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.
Methods: 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.
Results: 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/SIT124 ($n = 2/8$) and SIT3234 ($n = 8/8$) clustered together based on 24-locus MIRU-VNTR typing. The dominant sub-lineage was Beijing/SIT1 ($n = 19$), with the isoniazid resistance *katG* G944C mutation (Ser315Thr) detected in two of them.
Conclusions: The population structure of MTB in Kandy, Sri Lanka was different from that in the South Asian region. The clonal expansion of locally evolved lineage 4/SIT3234 and detection of the pre-multidrug resistant Beijing isolates from new tuberculosis patients is alarming and will require continuous monitoring.

Introduction

Tuberculosis (TB) is one of the oldest diseases known to mankind, yet it 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

* Corresponding author at: Division of Bioresources, Hokkaido University Research Center for Zoonosis Control, Kita 20, Nishi 10, Kita-ku, Sapporo, 001-0020, Japan.
** Corresponding author at: Department of Microbiology, Faculty of Medicine, University of Peradeniya, Sri Lanka.
E-mail addresses: medmicropera@pdn.ac.lk (C. Ratnatunga), cnakajim@czc.hokudai.ac.jp (C. Nakajima).

https://doi.org/10.1016/j.ijid.2019.07.001

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, respectively (World Health Organization, 2018).

It is believed that the emergence of multidrug-resistant TB (MDR-TB) and HIV, along with poor TB control, has contributed to the dramatic increase in the TB burden worldwide. According to the national surveillance conducted in 2018, the estimated percentage of TB cases with MDR-TB among new TB patients in Sri Lanka was 0.5%, while it was 4.1% among retreatment patients. Sri Lanka has a relatively good TB control program, 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 changes in the socio-cultural environment, an increasing prevalence of diabetes mellitus, and increased 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 that are more prone to develop drug resistance or are more transmissible and virulent. Hence monitoring the MTB population will provide important data to monitor and underpin the Sri Lankan TB control program.

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 subpopulations (Gagneux et al., 2006b). Although MTB shows a strong phylogeographical population structure, some lineages occur globally while others show a strong geographical restriction. For example lineages 2 and 4 are widespread globally, probably due to high virulence, compared to lineages 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). Understanding the genetic diversity of MTB strains in a given clinical setting is thus 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, providing a platform to study the genetic diversity, transmission dynamics, and phylogenetics 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 to lineages 1–6. Spoligotyping is a frequently used PCR-based molecular typing technique that 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). The MIRU-VNTR method has been used along with spoligotyping, as the combination of the two approaches has more discriminatory power to identify epidemiologically linked strains.

The molecular epidemiology of MTB is poorly explored in Sri Lanka. Although several studies that have applied molecular DNA fingerprinting techniques such as IS*6110* restriction fragment length polymorphism (RFLP), spoligotyping, and MIRU-VNTR have been performed (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 the circulating genotypes of MTB and their transmission patterns within the district of Kandy, in Central Province in Sri Lanka by using spoligotyping, LSP analysis, and MIRU-VNTR typing.

Materials and methods

Sample collection

Sputum samples were collected from 100 randomly selected 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 over the age of 18 years and currently residing in Kandy District were included in this study. The collected sputum samples were processed and cultured on Lowenstein–Jensen medium in the Department of Microbiology, Faculty of Medicine, University of Peradeniya. Data on patient demographics, risk factors, and laboratory investigations were also collected.

DNA extraction

Suspected MTB colonies grown on Lowenstein–Jensen medium were suspended in 200 μl of distilled water and heated for 20 min 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-associated genes

Comparative sequence analysis of the *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 described previously (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. The presence or absence of such spacer was determined and this pattern was converted into a 43-digit binary code system that 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 no assigned SIT or sub-lineage were analyzed by LSP to assign lineages. PCR was performed

using specific primers for the expected RDs, namely lineage 1-RD239 and 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* (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 did not 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

The statistical analysis was performed using RStudio (Integrated Development for R, RStudio, Inc., Boston, MA, USA; URL http://www.rstudio.com/). A spoligoforest tree (Fruchterman–Reingold algorithm) was drawn using spolTools online software (Reyes et al., 2008; Tang et al., 2008) available at http://spoltools.emi.unsw.edu.au/ to identify the evolutionary relationships 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-locus MIRU-VNTR pattern, and the clustering rate was calculated using the following 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 lineages 1 and 4, these were excluded. Finally, 85 isolates were used for molecular analysis. All suspected TB patients living in the district are supposed to visit the Central Chest Clinic. The estimated population of the district in 2012–2013 was around 1.37 million and the TB incidence rates in 2012 and 2013 were 46.6 and 52.2 per 100 000, respectively. Thus, the 89 samples can be taken as representative of the region and comprised approximately one in seven of the expected total TB incident cases in Kandy District during the collection period.

Table 1
Description of 26 spoligotype international types (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 with SIT number but the sub lineage is not assigned; New type: Spoligotype pattern found in our study

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* or QRDR in *gyrA*.

Spoligotyping and LSP

Spoligotyping of 85 isolates enabled the detection of 26 distinct spoligotype patterns, corresponding to 21 different SITs and five new patterns that have not been reported in the SITVIT2 database yet (Table 1). These new patterns (new types 1–5) were assigned to lineages 1 and 4 by LSP. The dominant lineage in this study was lineage 4 ($n = 39$, 45.9%), followed by lineage 1 ($n = 25$, 29.4%) and lineage 2 ($n = 20$, 23.5%). Only one isolate from lineage 3 was found (1.2%). The ratio of lineage 4 was significantly higher than that of the 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 (**Supplementary Material** Table S1).

MIRU-VNTR typing

Based on the spoligotyping results, clusters of Beijing/SIT1 + SIT190 ($n = 17/20$; three isolates were excluded when constructing the MST due to no bands in several loci), EAI3_IND/SIT11 ($n = 16$), undesignated lineage 4/SIT124 ($n = 8$), and undesignated lineage 4/SIT3234 ($n = 8$) were analyzed by 24-locus MIRU-VNTR typing and an MST was constructed (Figure 1). The clustering rate in the Beijing sub-lineage, SIT11, and SIT124 was 41%, 56%, and 50%, respectively. All eight isolates in SIT3234 were in one cluster (clustering rate = 100%) together with two isolates of SIT124. Genetically isoniazid-resistant isolates in SIT1 ($n = 2$) were singletons (Figure 1).

Analysis of patient demographics, risk factors, and laboratory findings

Complete data for patient demographics, risk factors, and laboratory findings (smear positivity and time to culture positivity) were available for 55 patients (**Supplementary Material** Table S1). Overall, 42 patients were male and 13 were female (male to female ratio, 3.23:1). The age of the patients ranged from 21 to 80 years. There was no significant association between variables and lineage 4 or non-lineage 4 in category-wise comparisons (Fisher's exact test, $p > 0.05$ in all categories).

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, lineages 1, 2, 3, and 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 the present study, only lineage 1 (labeled EAI or MANU

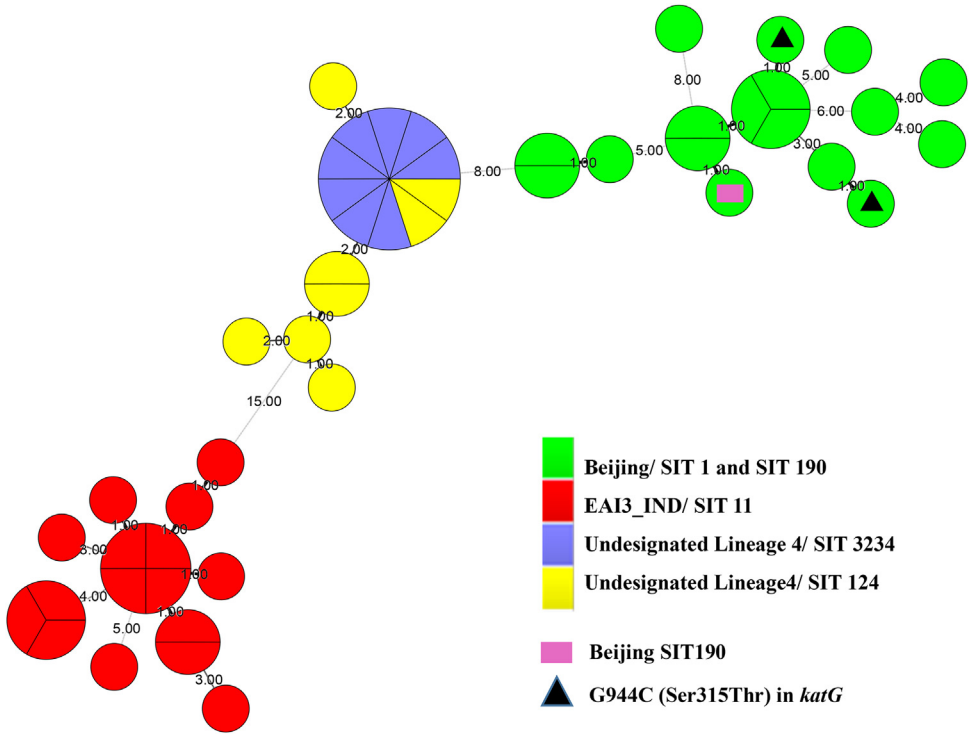

Figure 1. 24-Locus MIRU-VNTR-based minimum spanning tree of Beijing/SIT1 + SIT190, EAI3_IND/SIT11, undesignated lineage 4/SIT124, and undesignated lineage 4/SIT3234 isolates.
Each node represents a MIRU-VNTR type. The size of the node indicates the number of isolates in each cluster. The length of the branch represents the distance between patterns, while the number on the branch denotes the number of loci changes between two patterns. Green: Beijing sub-lineage; red: EAI3_IND/SIT11; yellow: undesignated lineage 4/SIT124; purple: undesignated lineage 4/SIT3234. Beijing/SIT190 is marked with a pink rectangle ($n = 1$). Isolates having the G944C (Ser315Thr) mutation in *katG* are marked with a black triangle ($n = 2$). (For interpretation of the references to colour in the figure legend, the reader is referred to the web version of this article.)

Table 2
Summary of *Mycobacterium tuberculosis* lineage distribution from previous studies in Sri Lanka and India.

Country of isolation	Sri Lanka				India			
	Ratio of each lineage (%)							
Reference	Current study	Rajapaksa et al. (2008)	Joseph et al. (2013)	Manson et al. (2017a)	Thomas et al. (2011)	Sharma et al. (2017)	Varma-Basil et al. (2016)	Gutierrez et al. (2006)
Sample number	N = 85	N = 98	N = 168	N = 201	N = 101	N = 335	N = 139	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[b]	Ghatampur[b]	Delhi[b]	Whole country
Lineages								
Lineage 1	29.4	58.2	81.5	70.0	48.5	22.4	23.0	45.0
Lineage 2	23.5	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	45.9	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, Tamil Nadu: Southern India; Andhra Pradesh: South Eastern India; Ghatampur, Delhi: Nothern India.

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 2, **Supplementary Material** Table S2).

In a previous study in Sri Lanka in which isolates were collected in Colombo, the 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 2, **Supplementary Material** Table S2), suggesting less interaction between Sri Lanka and central or northern India.

In contrast, the present study results revealed that the predominant lineage circulating in 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 (16th–17th, 17th–18th, and 19th-mid 20th century, respectively). It is hence hypothesized that the introduction of lineage 4 into Sri Lanka may have happened during the European colonial period. Supporting our hypothesis, population genomic and phylogeographic analyses of MTB lineage 4 have found that the 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 the predominance of lineage 4 in 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+), which may have been the endemic lineage in Kandy prior to colonization.

Fourteen distinct spoligotype patterns were identified in the lineage 4 isolates. Half of them were designated as Haarlem, T, and X sub-lineages, 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, the Netherlands, and the UK; SIT2 has mainly been 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 sub-lineages seem to be well established in Sri Lanka, as they have also been 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 these 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 microevolution to adapt to the Sri Lankan host population may have occurred. However, further detailed studies using techniques such as whole genome sequencing and time-scaled haplotypic density are warranted to confirm the factors that have shaped the local population structure of MTB in Sri Lanka.

A spoligoforest tree (Figure 2) revealed the probable parental links between the strains belonging to the different sub-lineages. Most of the 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 that all isolates ($n = 8$) in SIT3234 were in one cluster together with two isolates of SIT124, indicating a clonal expansion of these sub-lineages in the study group. Of the eight SIT3234 isolates, the patient demographic data were available for four, and these revealed that the patients all lived in different areas and that there was no direct contact between them, suggesting that this sub-lineage has already spread widely in the area. In the SITVIT2 database, 0.06% of isolates belong to SIT124, with a worldwide distribution that includes India, China, the Netherlands, and the UK, all of which are known to have deeply rooted historical relations with Sri Lanka. Previous studies have also suggested the clonal expansion of this sub-lineage 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, has also been reported in Sri Lanka (Weerasekara et al., 2015). The comparison of 15-locus MIRU-VNTR patterns of SIT124 and SIT3234 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 sub-lineage, 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 neighbor-joining 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' sub-lineages like Beijing and Haarlem are suggested to be more virulent compared to 'ancient' ones such as EAI. Based on this assertion, SIT124 and SIT3234, which showed clonal expansion in this study, could have implications for the epidemiology and control of TB in Sri Lanka in the future.

The Beijing sub-lineage is considered to be one of the predominant MTB sub-lineages, with a worldwide distribution and particularly dominating in the countries of East and Southeast Asia (Tamaru et al., 2012; Merker et al., 2015). The Beijing sub-lineage is suggested to be more virulent than other sub-lineages, 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

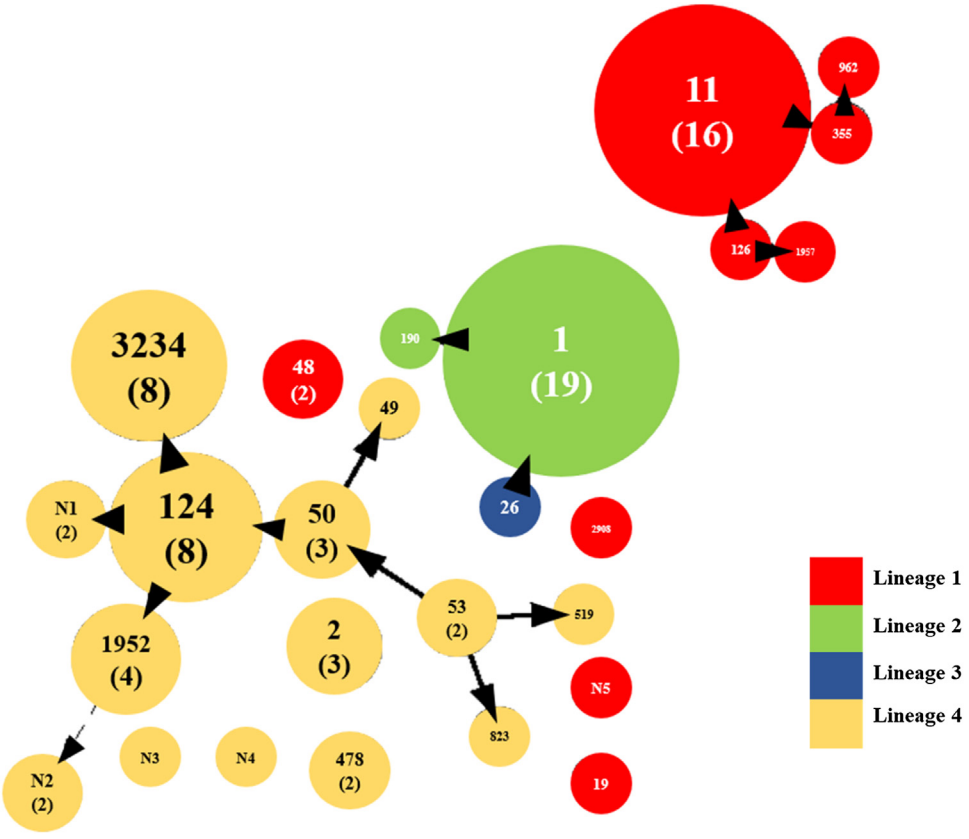

Figure 2. Spoligoforest tree based on all spoligotypes.
Each spoligotype pattern from the study is represented by a node, with the area size being proportional to the total number of isolates with that specific pattern. Changes (loss of spacers) are represented by directed edges between nodes, with the 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 weight being 1.0), while dashed lines represent links of weight between 0.5 and 1 and dotted lines a weight less than 0.5. The number inside the circle is the SIT number, while the number in parenthesis indicates the number of isolates in this study with that SIT.

sub-lineage responsible for MDR and XDR (Parwati et al., 2010; Merker et al., 2015; San et al., 2018). Ongoing transmission of the Beijing sub-lineage has previously been detected in Sri Lanka (Rajapaksa et al., 2008; Weerasekera et al., 2015), as well as in the current study. While SIT1 was the most prevalent SIT that we found, MIRU-VNTR results (Figure 1) showed highly diverse patterns. The Beijing lineage may have been introduced to Sri Lanka through trading links with Southeast Asian countries during the period that Sri Lanka was one of the main ports in ancient maritime silk and spice trade routes. Furthermore, the continuous migration and emigration between populations in Sri Lanka, China, and other South Asian countries that continues up to the present day may be responsible for the higher genetic diversity within this sub-lineage 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 that may also need to be considered for selective transmission of Beijing strains in Sri Lanka, as there is high coverage of BCG vaccination.

Two isolates from the Beijing/SIT1 clade had a G944C mutation (Ser315Thr) in *katG*, suggesting resistance 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 resistance 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 mutation, *katG* Ser315Thr, also known as the pre-MDR-TB mutation, could be advantageous in this respect.

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), suggesting that South India may represent the probable origin of this sub-lineage in Sri Lanka due to migratory patterns that stretch back to ancient times. Furthermore, in a previous study by Rajapaksa et al. (2008), the EAI5 sub-lineage was shown to be prevalent ($n = 20/98$, 20%) in the Western Province, while present at a much lower prevalence in Kandy District ($n = 3/85$, 3.5%). These findings suggest the diversity of the MTB population structure in Sri Lanka. A high proportion of the MANU sub-lineage was detected in Kandy by Weerasekera et al. (2015), but we were unable to identify any isolate within this sub-lineage. This discrepancy may have occurred because the MANU sub-lineage 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 period of time with microevolution driving the emergence of new descendants, which may have adapted to the local Sri Lankan host population. Therefore, the clonal expansion of locally evolved and potentially host-adapted undesignated lineage 4/SIT3234 requires continuous 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. Although no MDR-TB was found in this study, two isolates of Beijing /SIT1 from new TB patients had the well-known pre-MDR *katG* G944C mutation (Ser315Thr), which warns of the need for monitoring. This study shows that it will be necessary to conduct 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.

Funding

This study was supported by the Japan Agency for Medical Research and Development (AMED) under grant numbers JP18fm0108008, JP18fk0108042, and JP18jk0210005, and partially supported by a grant from the Ministry of Education, Culture, Sports, Science and Technology, Japan (MEXT) for the Joint Research Program of the Hokkaido University Research Center for Zoonosis Control.

Ethics statement

This study received ethical approval from the Ethics Review Committee, Faculty of Medicine, University of Peradeniya, Sri Lanka.

Conflict of interest

None.

Acknowledgements

We would like to thank the staff of the Central Chest Clinic, Kandy, Sri Lanka and Department of Microbiology, Faculty of Medicine, University of Peradeniya, Sri Lanka for helping with the collection and processing of samples, and the staff of Hokkaido University Research Center for Zoonosis Control, Japan for their support.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.ijid.2019.07.001.

References

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.

Brosch R, Gordon SV, Marmiesse M, Brodin P, Buchrieser C, Eiglmeier K, et al. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. Proc Natl Acad Sci U S A 2002;99(6):3684–9.

Brynildsrud OB, Pepperell CS, Suffys P, Grandjean L, Monteserin J, Debech N, et al. Global expansion of *Mycobacterium tuberculosis* lineage 4 shaped by colonial migration and local adaptation. Sci Adv 2018;4(10)eaat5869.

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

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 2019;, doi:http://dx.doi.org/10.1016/j.meegid.2018.12.030.

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 2016;96:13–20, doi:http://dx.doi.org/10.1016/j.tube.2015.11.002.

Gagneux S, Burgos MV, 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 2006b;103(8):2869–73, doi:http://dx.doi.org/10.1073/pnas.0511240103.

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

Glynn JR, Vynnycky E, Fine PE. Influence of sampling on estimates of clustering and recent transmission of *Mycobacterium tuberculosis* derived from DNA finger-printing techniques. Am J Epidemiol 1999;149:366–71.

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 2006;12(9):1367–74.

Hall T. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 1999;41:95–8.

Helb D, Jones M, Story E, Boehme C, Wallace E, Ho K, et al. Rapid detection of *Mycobacterium tuberculosis* and rifampin resistance by use of on-demand, near-patient technology. J Clin Microbiol 2010;48(1):229–37, doi:http://dx.doi.org/10.1128/JCM.01463-09.

Joseph BV, Soman S, Radhakrishnan I, Hill V, Dhanasooraj D, Kumar RA, et al. Molecular epidemiology of *Mycobacterium tuberculosis* isolates from Kerala, India using IS*6110*-RFLP, spoligotyping and MIRU-VNTRs. Infect Genet Evol. 2013;16:157–64, doi: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 *Mycobacterium tuberculosis* for diagnosis and epidemiology. 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 2016;5:7–10, doi:http://dx.doi.org/10.1016/j.jgar.2016.01.007.

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 2012;12(4):798–806, doi:http://dx.doi.org/10.1016/j.meegid.2011.08.028.

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

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, doi:http://dx.doi.org/10.1093/cid/cix169.2017.

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. 2017b;49(3):395–402, doi: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 2015;47(3):242–9, doi:http://dx.doi.org/10.1038/ng.3195.

O'Neill MB, Shockey A, Zarley A, Aylward W, Eldholm V, Kitchen A, et al. Lineage specific histories of *Mycobacterium tuberculosis* dispersal in Africa and Eurasia. Mol Ecol 2019;, doi:http://dx.doi.org/10.1111/mec.15120 [in print].

Parwati I, van Crevel R, van Soolingen D. Possible underlying mechanisms for successful emergence of the *Mycobacterium tuberculosis* Beijing genotype strains. Lancet Infect Dis 2010;10(2):103–11, doi: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.

Rajapaksa US, Victor TC, Perera AJ, Warren RM, Senevirathne SM. Molecular diversity of *Mycobacterium tuberculosis* isolates from patients with pulmonary tuberculosis in Sri Lanka. Trans R Soc Trop Med Hyg 2008;102(October (10)):997–1002, doi:http://dx.doi.org/10.1016/j.trstmh.2008.04.025.

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

San LL, Aye KS, Oo NAT, Shwe MM, Fukushima Y, Gordon SV, et al. Insight into multidrug-resistant Beijing genotype *Mycobacterium tuberculosis* isolates in Myanmar. Int J Infect Dis. 2018;76:109–19, doi:http://dx.doi.org/10.1016/j.ijid.2018.06.009.

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 2017;63:13–20, doi: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 (Edinb) 2017;105:60–72, doi:http://dx.doi.org/10.1016/j.tube.2017.04.003.

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, 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.

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, Jamnani FR, Siadat SD, Vaziri F. Prevalence of Beijing and Haarlem genotypes among multidrug-resistant *Mycobacterium tuberculosis* in Iran: systematic review and meta-analysis. Tuberculosis 2017;107:31–7, doi:http://dx.doi.org/10.1016/j.tube.2017.03.005.

Thomas SK, Iravatham CC, Moni BH, Kumar A, Archana BV, Majid M, et al. Modern and ancestral genotypes of *Mycobacterium tuberculosis* from Andhra Pradesh, India. PLoS One 2011;6(11)e27584, doi:http://dx.doi.org/10.1371/journal.pone.0027584.

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 U S A 2004;101(14):4865–70.

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 2016;100(September):72–81, doi:http://dx.doi.org/10.1016/j.tube.2016.07.007.

Weerasekera D, Magana-arachchi D, Madegedara D, Dissanayake N, Thevanesam V. Genetic diversity of *Mycobacterium tuberculosis* isolates obtained from three distinct population groups in the Central Province, Sri Lanka. Asian Pac J Trop Dis 2015;5(5):385–92, doi: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 (Lond) 2019;51(3):206–15, doi:http://dx.doi.org/10.1080/23744235.2018.1551619.

Weniger T, Krawczyk J, Supply P, Niemann S, Harmsen D. MIRU-VNTRplus: a web tool for polyphasic genotyping of *Mycobacterium tuberculosis* complex bacteria. Nucleic Acids Res. 2010;38 Suppl:W326–331.

World Health Organization. Global tuberculosis report. 2018 WHO/CDS/TB/2018.20. ISBN 978-92-4-156564-6.