IS*6110* FAFLP PCR, a tool for genomic mapping enabling investigation of evolutionary relationships of *Mycobacterium tuberculosis* in resource poor settings

Kartykayan Moganeradj

Doctor of Philosophy

UCL

Declaration

I, Kartykayan Moganeradj, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed:

Abstract

Tuberculosis (TB) is an important communicable disease affecting the human population world-wide. Despite the efforts of the scientific community, national governments and WHO in controlling the disease, TB still remains a major killer in resource poor settings. New rapid assays and techniques that are simple and cost-effective are urgently needed to identify, treat and understand pathogenesis including the geographical distribution of the disease. The aim of the thesis is to develop a novel genomic mapping tool using Insertion Element, IS*6110* that could aid in epidemiological studies of *Mycobacterium tuberculosis* complex (MTBC) in low and middle income countries.

IS*6110*, a bacterial transposon, plays an essential role in changing the physical and biochemical traits of MTBC. Due to their transposition in TB genomes, they are used as epidemiological markers for differentiation of TB organisms and the mapping of these elements could also shed light on the putative altered function of adjacent genes. In the era of Whole Genome Sequencing (WGS) where repeat elements are difficult to sequence with short read technologies, a rapid and simple method of insertion site mapping using IS*6110* FAFLP PCR was developed. This work is aimed at developing a rapid, cost-effective and robust genomic tool box exploiting the IS*6110* FAFLP PCR assay that can both identify and characterise the TB genotypes / genetic lineages in any geographical location.

For the first time using the assay above, TB samples from Nepal were categorised into different genetic lineages. Fifty-five percent of the samples analysed belong to Principal Genetic Group 1 (PGG1), Beijing and Central Asian strains. Also, new primers were designed targeting the Beijing and the T- groups using the FAFLP derived data that gave rise to the development of rapid lineage specific PCR assays. In addition, it was noticed that 3.9% of the Nepalese strains tested in this research work were likely multi-drug resistant (MDR-TB) using PCR targeting the Rifampicin-resistance-determining region (RRDR) of the rpoB region.

It is demonstrated here that IS*6110* FAFLP methodology could easily characterise the TB samples into different genetic lineages provided they have more than four IS*6110* copies. In addition, lineage specific PCR does not need any expensive instruments or reagents except for PCR blocks and gel visualisers, and could be very effective in the rapid identification of different TB genotypes within hours. These data also add to knowledge about the circulating strains of TB in Nepal, currently a poorly characterised region of the world in this regard, and could help in contact tracing studies by epidemiologists. The IS*6110* FAFLP technique thus can be employed in any geographical location to map TB genetic lineages where there is little or no information available on the prevailing TB strains.

Table of Contents

List of Tables

List of figures

Abbreviations

AFLP- Amplified Fragment Length Polymorphism

CAS- Central Asian Strain

DOTS- Directly Observed Treatment short-course

DR- Direct Repeat

DR-TB- Drug Resistant Tuberculosis

DST – Drug Susceptibility Test

DVR- Direct Variable Repeats

ETH- Ethambutol

FAFLP- Fluorescent Amplified Fragment Length Polymorphism

GENETUP- German Nepal Tuberculosis Project

HIV- Human Immunodeficiency Virus

IE- Insertion Element

INH- Isoniazid

ipl- IS*6110* preferential loci

IS- Insertion Sequence

LAM- Latin American- Mediterranean

MDG- Millennium Development Goal

MDR- Multidrug Resistant

MIRU– Mycobacterial Interspersed Repetitive Units

MPTR- Major Polymorphic Tandem Repeats

MTB- *Mycobacterium tuberculosis*

MTBC- *Mycobacterium tuberculosis* Complex

NTP- National Tuberculosis Programme

PCR- Polymerase Chain Reaction

PRZ- Pyrazinamide

RFLP- Restriction Fragment Length Polymorphism

RIF- Rifampicin

TB- Tuberculosis

TDR- Totally Drug Resistant

VNTR- Variable Number Tandem Repeats

WHO- World Health Organisation

XDR- Extensively Drug Resistant

Glossary of Key terms

Genotype- Genetic makeup of any organism such as *Mycobacterium tuberculosis*.

MDR-TB- Multidrug-resistant TB (MDR-TB) is TB that is resistant to at least isoniazid and rifampicin, the two most powerful anti-TB drugs.

PCR- Polymerase chain reaction is an enzymatic method where one copy of segment of DNA is amplified into several million copies of DNA.

Tuberculosis- Infectious disease caused by the bacteria *Mycobacterium tuberculosis* complex

XDR-TB- Form of TB which is resistant to at least four of the core anti-TB drugs including isoniazid and rifampicin.

Outline of the thesis

This research work is divided into chapters that explain the development of an IS*6110* FAFLP genomic tool box that could be used in resource poor settings.

Chapter 1 sets the scene for this thesis by reviewing the literature extensively. It gives a brief introduction about tuberculosis infection including diagnostics, treatment and TB epidemiology. Then an extensive review of the available molecular epidemiological markers is discussed including the IS*6110* marker. The chapter ends with the overview of the study setting (Nepal), leading to aims and objectives of this research work.

Chapter 2 discusses the general methods that were used to achieve the objectives of this project.

Chapter 3 describes in detail how the IS*6110* FAFLP PCR assay was developed and standardised.

Chapter 4 shows for the first time how the standardised method was tested by mapping the IS*6110* insertion sites in the reference genome H37Rv and by showing the development of rapid detection PCR targeting specific lineages.

Chapter 5 describes characterisation of TB samples from Nepal, a resource poor setting, into different TB genetic lineages for the first time using IS*6110* FAFLP PCR.

Chapter 6 describes rifampicin sensitivity in the Nepalese TB samples using RIF resistant PCR assay.

Chapter 8 finally summarises all the above chapters and concludes by putting the results into context.

My Role in the thesis

I started on this research work in 2011 as a part-time PhD student whilst still continuing my day job as a Specialist Healthcare Scientist in PHE, Colindale. I was supervised by my manager, Dr. Catherine Arnold, who was also my principal supervisor for my thesis work. My three other supervisors were Prof. Tim McHugh (UCL), Prof. Pam Sonnenberg (UCL) and Prof. Ibrahim Abubakar (UCL).

Under the able supervision of my principal supervisor and the guidance of my other supervisors, I designed and developed the method called IS*6110* FAFLP PCR. I tested the procedure on the Nepal TB samples donated kindly by our collaborators in Nepal, Dr. Deanna Hagge and Ms. Saraswoti Khadge.

I initially mapped the insertion sites in the MTB genome H37Rv followed by designing new two-base TB genotype/ lineage specific primers that could rapidly identify and characterise the circulating lineages in any geographical location. At the same time, I also characterised the Nepalese TB samples into different TB lineages. At the time of this experiment, it was the first known experiment to delineate Nepalese TB samples as there was no information available publicly on the distribution of the circulating lineages in this region. During the course of the thesis, I also tested the drug resistance status of these Nepalese strains.

During the course of the thesis, I have published three manuscripts in peer reviewed journals (as discussed in chapters 4, 5 and 6). I have also presented a number of posters in various domestic and international conferences. I have also attended the skills courses both at UCL and PHE and updated my skills during the thesis.

Publications and conference presentations

Manuscripts published from this research work

- Moganeradj K, Abubakar I, Mchugh TD, Sonnenberg P, Arnold C (2013) Insertion site mapping for repeated elements in *Mycobacterium tuberculosis*. J Microbiol Methods 92: 192–196.

- Moganeradj K, Rajendram D, Khadge S, Sonnenberg P, Abubakar I, et al. (2016) *Insertion element IS6110 based characterisation of Nepalese tuberculosis strains into different genetic lineages*. Clin Microbiol Infect Dis, 1.

- Hameed S, Moganeradj K, Mahmood N, McHugh TD, Chaudhry MN, Arnold C (2017) Sequence analysis of the rifampicin resistance determining region (RRDR) of rpoB gene in multidrug resistance confirmed and newly diagnosed tuberculosis patients of Punjab, Pakistan. PLoS One, 12(8): e0183363. (This part of the work was carried out whilst I supervised another PhD student from Pakistan, by employing the tools used in chapter 6).

Posters presented

1. Kartyk Moganeradj, Dunstan Rajendram, Ibrahim Abubakar, Timothy D McHugh, Pam Sonnenberg, Deanna A Hagge, Murdo Macdonald, Saraswoti

Khadge and Cath Arnold (2016) *Insertion Sequence IS6110 mapping, a tool to characterise TB strains into genetic lineages.* ICID 2016- Hyderabad, India.

2. Kartyk Moganeradj, Dunstan Rajendram, Ibrahim Abubakar, Timothy D McHugh, Pam Sonnenberg, Deanna A Hagge, Murdo Macdonald, Saraswoti Khadge and Cath Arnold (2014) *Low Burden of Rifampicin Resistant TB Lineages in Nepal.* ECCMID 2014- Barcelona, Spain.

3. Kartyk Moganeradj, Ibrahim Abubakar, Timothy D McHugh, Pam Sonnenberg, and Cath Arnold (2013). *Rapid detection of TB genetic lineages using specific primers.* ECCMID 2013- Berlin, Germany.

4. Kartyk Moganeradj, Pam Sonnenberg, Ibrahim Abubakar, Tim McHugh, Deanna A. Hagge, Saraswoti. Khadge, Cath Arnold (2012) *Delineating TB samples from Nepal using Insertion Site 6110 (IS6110) mapping PCR.* ECCMID 2012- London, U.K.

Acknowledgements

I would like to thank all my four supervisors, Dr. Cath Arnold (PHE), Prof. Tim Mchugh (UCL), Prof. Pam Sonnenberg (UCL) and Prof. Ibrahim Abubakar (UCL) for their valuable time and overwhelming support, sharing their knowledge and experience in shaping my PhD thesis into a challenging and commendable piece of research work. Without their supervision and their unified guidance, I would not have completed this research and published three articles in peer-reviewed journals. My special thanks to Dr. Cath Arnold, my principal supervisor, for guiding and encouraging me throughout my PhD tenure, for pulling my spirits up whenever I was down and also helping me to achieve my goal of finishing the PhD. Her immense support by allowing me to work flexible hours when I was facing personal issues needs a special mention as I would have otherwise floundered during the middle of my research work.

I owe my thanks to my secondary supervisor, Prof. Tim McHugh, for being those extra pair of eyes in looking into my thesis and suggesting corrections. The numerous meetings I had with you definitely helped me to increase the level of scientific acumen needed for the research work.

Many thanks to my tertiary supervisor, Prof. Pam Sonnenberg, whose keen eye helped me to re-organise my thoughts and pay attention to the structure of the thesis.

I would like to thank my tertiary supervisor, Prof. Ibrahim Abubakar, whose constant encouragement motivated me whenever I was looking down the barrel during the PhD meetings.

I also thank my colleagues, both past and present, in the Genomics Services Development Unit, where I did my PhD part-time for giving me space and time to concentrate on my research work. I owe my special thanks to Dunstan, Ali, Des, Meeta, Kirstin, Julie, Raju, Danny, Sandra, Jumie, Greg and Andrew for their constant nagging to get my thesis finished. I also would like to extend my gratitude to my mentors, teachers and all my friends from my previous work places in U.K. and in India.

I take this opportunity to thank Public Health England (previously Health Protection Agency) for partly funding my PhD. I also thank University College London for accepting me as a PhD student part-time.

I would also like to thank my past manager from SGUL, Prof. Steve Goodbourn and his group, my RVC supervisor, Prof. Peter Russell and my RIVER professors especially Dr. S. Ram Kumar for supporting me at various stages of my career.

I owe my special thanks with a lot of gratitude to my parents and sister, who saw in me that one day I will definitely make them proud. I thank my aunty, Bibi and my guardians, Babu and Asha, for kick-starting my aspirations in U.K.

I thank overwhelmingly my wife, Vimala, for her support and holding hands with me firmly during this arduous but successful seven year journey and never letting me go mad.

Last by no means the least, I owe my special thanks to my angel and my beautiful daughter Yashwini who taught me how to stay strong and put up a smile, when faced with adversity. She always found it funny to see her Daddy working harder than her and studying at night.

I dedicate this thesis to my parents, my wife and my daughter who sacrificed a lot to make this happen for me.

Chapter 1 Introduction

Tuberculosis (TB) caused by the *Mycobacterium tuberculosis* complex (MTBC), is ranked as one of the top ten causes of death worldwide (WHO 2015) with an estimated 10.4 million people falling ill with TB and 1.3 million TB deaths in 2016, exacerbated by HIV co-infection and the development of drug resistance (1).

1.1 ***Mycobacterium tuberculosis*** **complex (MTBC)**

Mycobacterium tuberculosis belongs to the genus *Mycobacterium* in the family Mycobacteriaceae, order Corynebacteriales and phylum Actinobacteria (2).

TB caused by the MTBC encompasses the following members including proposed new members that affect mongooses and meerkats (3–6).

1. *M. tuberculosis*
2. *M. africanum* (phylogenetic variant of *M. tuberculosis*).
3. *M. bovis*
4. *M. caprae*
5. *M. microti*
6. *M. pinnipedii*
7. *“M. mungi”*
8. *“M. orygis”*

9. “M. *suricattae”*

10. *“M. canett*i”

1.2 TB Timeline

Before Robert Koch, two French physicians Rene Theophile Hyacinthe Laennec and Jean-Antoine Villemin reported in 1821 and 1868 respectively the first accounts of tuberculosis, even though the causative organism was unidentified at that time (7,8). However, Robert Koch’s enormous body of work from identification of the microorganism, developing the staining technique and culture methods for *Mycobacterium tuberculosis*, illustrating the mode of transmission to recommending isolation of the patients with active TB has to be considered as the greatest achievement in the science of TB and has now paved the way for various studies. Koch’s failed study to treat tuberculosis by using supernatants from *M.tuberculosis* culture which is now known as ‘tuberculin’ when injected in the skin gave rise to cell-mediated immune responses which are responsible for the clinical manifestations of the disease (9,10). Subsequently in the years that followed, various scientists and physicians were able to relate that the host immune responses especially cell-mediated immune responses played a major role in the pathogenesis of TB (9). The important discoveries and outcomes in relation to the TB timeline is given in the table 1.1 (adapted from (11)).

Table 1-1 TB timeline showcasing the important discoveries and events from the 19th century to the 21st century

Rober Koch discovers the causative organism of tuberculosis,*M. tuberculosis*	**1882**	**Koch's discovery**
An attenuated vaccine from *M.bovis* developed by French scientists Albert Calmette and Camille Guerin was put into use in humans	**1921**	**BCG vaccine**
Streptomycin discovered in 1943 by american scientists, Selman Waksman, Albert Schatz, and Elizabeth Bugie, was found to be exhibiting antibacterial properties along with para- amino salicylic acid (PAS)	**1943**	**Streptomycin (S)**
First successful use of streptomycin in trials in humans	**1948**	**Randomised Control Trials (RCT)**
Combination therapy of S+PAS+H for 24 months	**1952**	**Isoniazid's (H) use in treatment**
Pyrazinamide discovery	**1954**	**Pyrazinamide (Z)**
Rifampicin discovery	**1957**	**Rifampicin (R)**
Ethambutol discovery and replaces PAS in combination therapy for 18 months	**1961**	**Ethambutol [E]**
Rifampicin included in combination therapy for 9-12 months and first outbreak of drug resistant TB in US	**1970-**	**S+H+R+E therapy and drug resistance**
BCG vaccination included in UNICEF's expanded immunisation programme	**1974**	**UNICEF vaccination programme**
Short term course for 6-8 months was introduced	**1980-**	**H+R+Z+E**
First meeting by WHO to discuss TB-HIV coinfection	**1989**	**HIV impact**
WHO declares TB as major public health issue as the TB related mortality was higher than previous years	**1991**	**Global Impact**
Anti-TB drug resistance survey and a new framework for TB control including DOTS	**1995**	**DOTS implementation**
First genome sequencing of Mycobacterium tuberculosis	**1999**	**TB genome**
WHO framework to completely eliminate TB by 2050 by adopting STOP TB strategy globally	**2014**	**WHO initiative for TB elimination**
	2015	**STOP TB framework**
	2020	**Post 2015 global TB strategy**
	2050	**TB elimination**

(BCG- Bacille Calmette Guerin, UNICEF- United Nations Children's fund, WHO- World Health Organisation, DOTS- Directly Observed treatment short-course)

According to WHO in 2016, the Millennium Development Goal (MDG) by the United Nations to stop the spread of TB by 2015 has already been achieved in all six WHO regions, and in sixteen of the 22 high burden countries that collectively accounts for more than 80% of the TB cases worldwide. TB incidence has dropped to a level by 1.5 % every year since 2000 and consequently, it is 18% lower than the levels of 2000. However, it is still a major global health problem of the developing world in heavily populated countries like India and China which account for more than a quarter of new TB cases in addition to sub- Saharan Africa. The number of incident TB cases varies from country to country and it is predominantly lower in high-income countries. (fig.1.1) (12). The MDGs have now been replaced by Sustainable Development Goals (SDGs) and the Stop TB strategy to the End TB strategy as shown in the figure 1-2

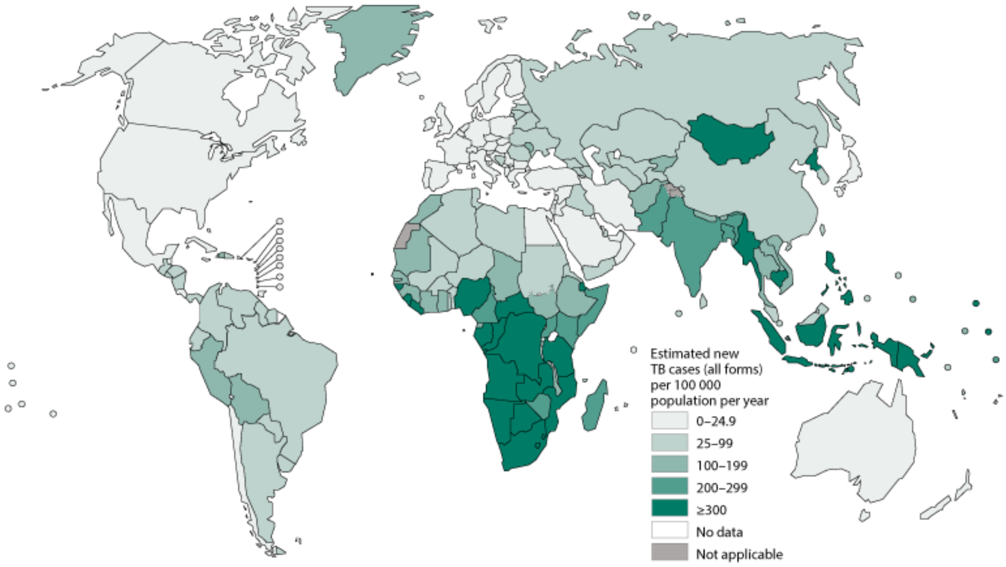


Figure 1-1 Estimated rates of incidence of new TB cases world-wide, 2015

(Taken from WHO’s global tuberculosis report 2016).

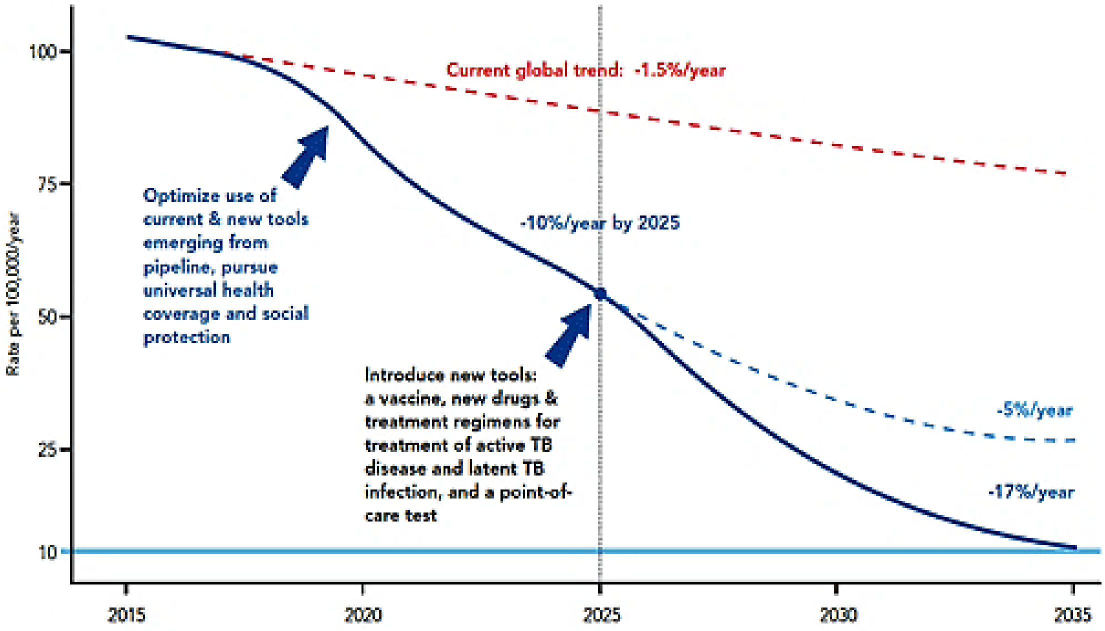


Figure 1-2 Milestones set by WHO to control TB by adopting End TB strategy

(Taken from http://www.who.int/tb/strategy/end-tb/en/)

1.3 Morphology of *Mycobacterium tuberculosis*

M. tuberculosis is a rod-shaped non-spore forming aerobic acid-fast bacterium (see figure 1-3) having a unique cell wall structure that supports their survival (13).

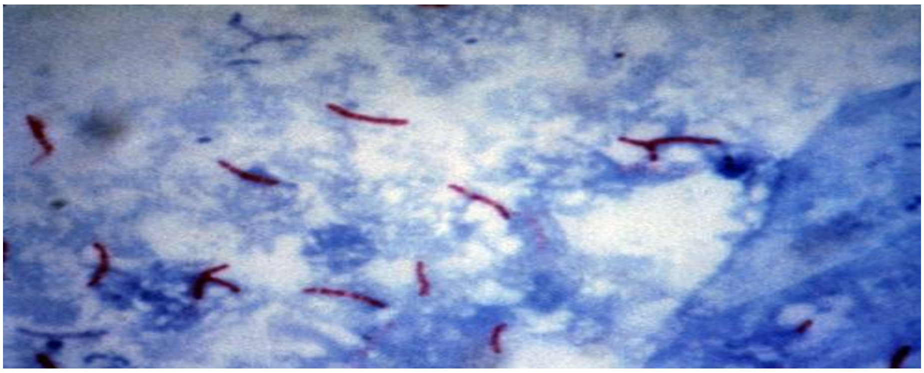


Figure 1-3 Acid fast Ziehl Neelsen staining showing rod-shaped *Mycobacterium tuberculosis* bacteria

(https://s-media-cacheak0.pinimg.com/originals/94/40/2d/94402db464e9849cd5aeb723cb90c91c.jpg)

The cell wall consists of mycolic acid covalently attached to the underlying peptidoglycan bound polysaccharide arabinogalactan providing a protective barrier against host defence mechanisms (fig 1-4) thus enabling the bacterium to be resistant to the action of antibiotics. In addition, the cell wall also contains lipoarabinomannan, a carbohydrate structural antigen on the outside of the organism that is immunogenic and facilitates the survival of mycobacteria within macrophages (14,15).

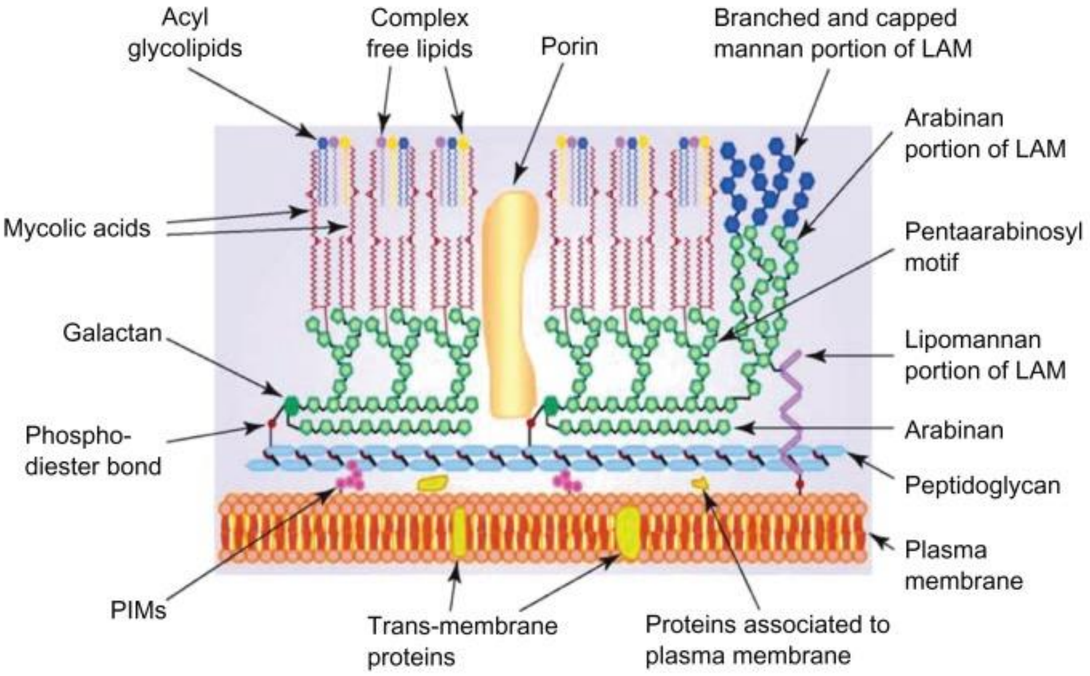


Figure 1-4 Unique Cell wall structure of *Mycobacterium tuberculosis*

(Taken from http://www.cell.com/cms/attachment/607349/4834477/gr1.jpg)

1.4 Immunopathogenesis of Tuberculosis

M. tuberculosis infection is caused by aerosols generated from a person infected with pulmonary or laryngeal tuberculosis. These droplets containing two to three *M. tuberculosis* organisms can stay airborne for minutes to hours after expectoration (16). The size (1-5 µm in range) and concentration of the aerosols also have been suggested to increase the TB transmission risk in patients with intra thoracic disease (17). The number of bacilli in the droplets, the virulence of the bacilli, exposure of the bacilli to the UV light, degree of ventilation and occasions for aerosolisation all influence transmission (fig. 1-5). *M. tuberculosis* not only affects the respiratory system when it enters the lungs but can also affect the lymphatic system, pleura, bones/joints, or meninges and thus cause extra pulmonary tuberculosis (14).

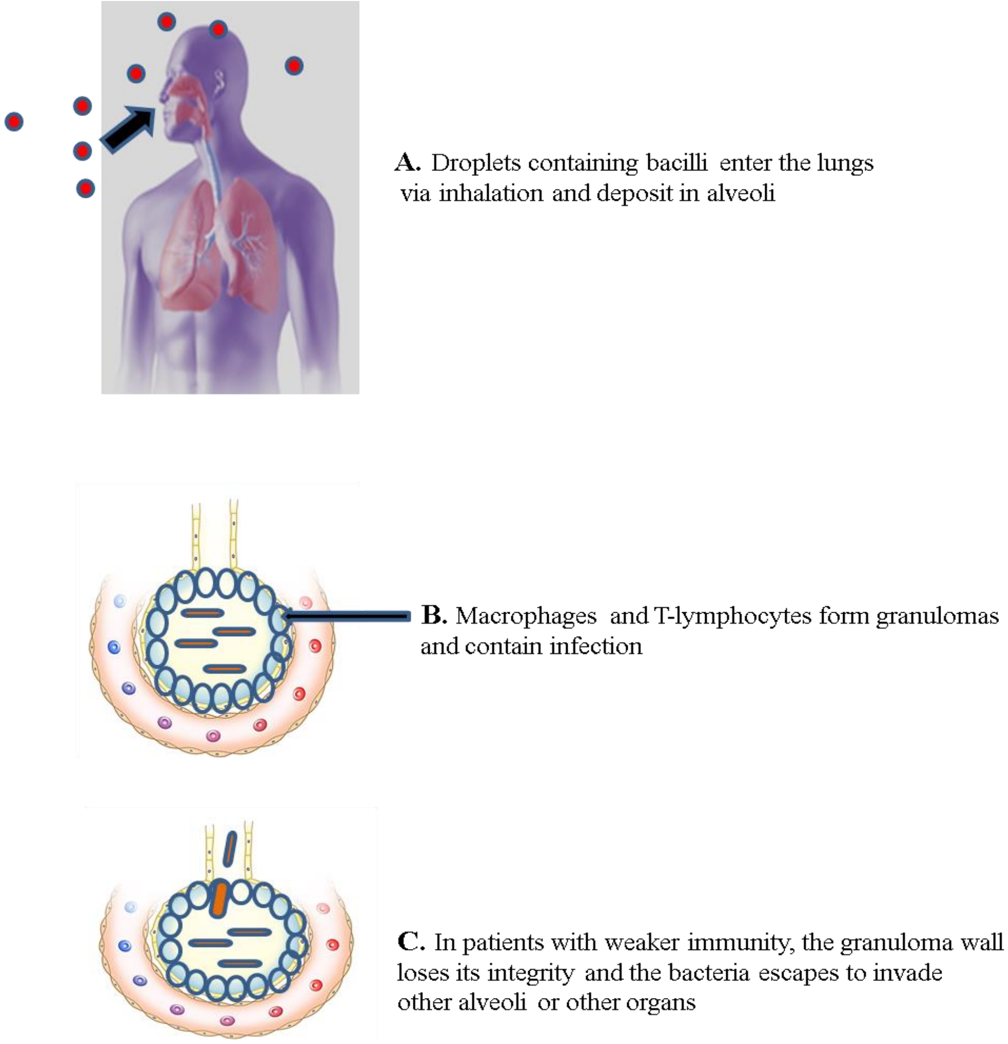


Figure 1-5 Pathophysiology of Tuberculosis infection

A) Inhalation of bacilli, B) Formation of granulomas containing Macrophages and T-lymphocytes and C) Collapse of granuloma and release of alveoli to other alveoli or other organs (Adapted from CDCP).

1.5 Clinical signs and symptoms

The active lung tuberculosis is characterised by cough with sputum (blood at times), chest pains, loss of appetite, general weakness, weight loss, fever, and night sweats.

The various stages of disease presentation in TB are as follows (13) as shown in figure 1-5:

- In early infection or primary disease, the hosts' immune system combats the infection and the patients present with fever, paratracheal lymphadenopathy or dyspnoea. Infection may not progress to active disease and could be subclinical
- There are two stages in the active disease presentation. In early primary progressive stage or active disease, the immune system fails to prevent the onset of infection and the inflammation of tissues follows. As the patients have nonspecific signs like fatigue, weight loss and fever followed by the development of non-productive cough, the diagnosis could be difficult: It is because the findings on chest radiographs might be normal and the sputum smear microscopy might not contain any mycobacteria ('negative' result).
- In late primary progressive/ active disease or active tuberculosis, the symptoms seen are productive cough leading to purulent sputum, progressive weight loss, chills, night sweats, respiratory crackles and anaemia. This state is an infectious state harbouring more bacilli than latent infection (18). Due to the excessive loss of both fat and lean tissue, this disease is also called wasting disease. Diagnosis is carried

out by sputum cultures as sometimes the chest radiographs appear normal.

- In latent infections or latent tuberculosis infection (LTBI), mycobacteria persist in the body without any clinical signs or symptoms and the disease could be reactivated when the immune system fails to counteract the infection or if immunosuppressed. Also, calcification and fibrosis of the granulomatous lesions are noticed in radiological findings. The formation of granulomas for persons with intact cell-mediated immunity is an accumulation of activated T-lymphocytes and macrophages. This micro-environment destroys the macrophages and produces early solid necrosis at the centre of the lesion but the bacteria survive this condition by modifying its phenotypic expression as suggested by Li et al. (19), by altering its expression to changes to pH and anaerobic growth conditions. By two or three weeks, the necrotic environment resembles soft cheese called caseous necrosis characterised by low pH, low oxygen levels and limited nutrients which leads to latency. Persons with adequate immunity develop fibrosis and calcification which contains the bacilli in these lesions whereas in those with impaired immunity develop primary progressive TB (20). The necrotic material undergoes liquefaction and the fibrous wall loses structural integrity in less immunocompetent persons which can then drain into the bronchus or a nearby blood vessel. Unique histopathological features of the lungs (granulomas of multinuclear cells, giant cells and caseation) in pulmonary tuberculosis are seen in figure 1-6. If the infected person coughs, droplet infection can occur and also if discharged into the blood vessel it leads to extra pulmonary tuberculosis (fig 1.5).

- In extra pulmonary tuberculosis, if it affects the central nervous system it leads to meningitis which left untreated results in deaths. If it affects the circulatory system causing infection of the blood stream, it leads to disseminated or miliary tuberculosis. When it affects the lymphatic system, it is called lymphatic tuberculosis and the most common presentation is cervical lymphadenopathy. In some cases, it is also shown to affect the joints, bones, pleura and the urogenital system (13).

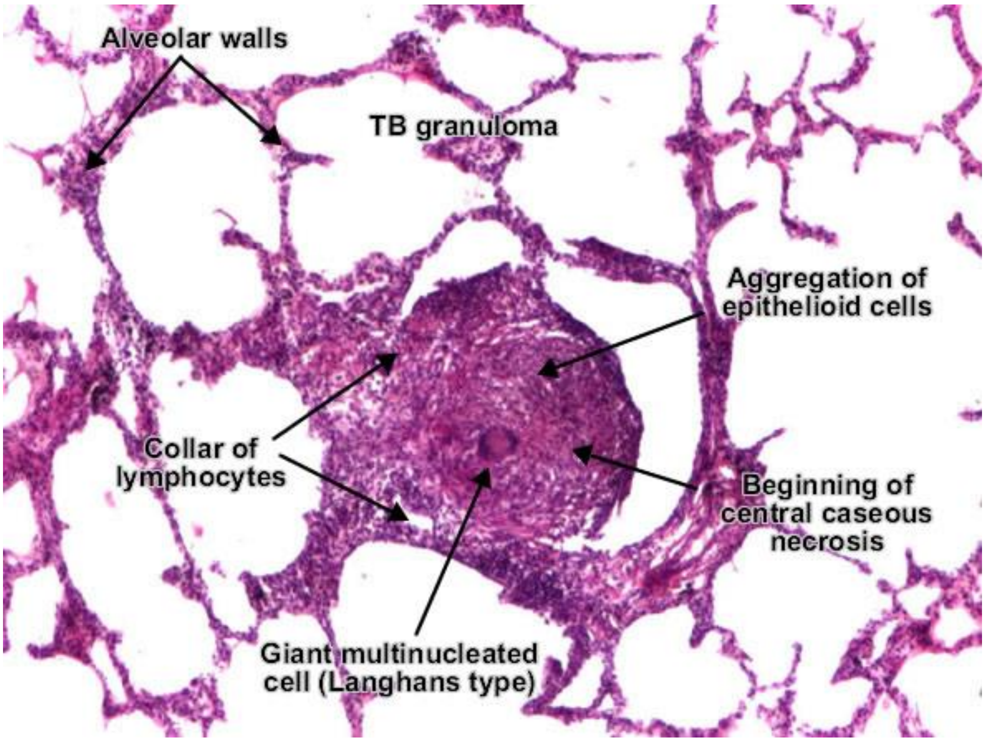


Figure 1-6 Histopathological features of lung infected by pulmonary tuberculosis showing caseation and formation of granulomas

(Accessed from http://intellectualventureslab.com/assets_uploads/pulmonary_diagram_inflammation.jpg)

1.6 Diagnosis of active tuberculosis

For TB treatment to be effective, a rapid and accurate diagnosis of drug resistant TB is important in ensuring the timely clinical intervention and controlling the rise in drug resistant TB (21,22). According to the WHO Global Tuberculosis report 2014, only 64% of the estimated nine million people were reported for newly diagnosed TB cases and more than three million cases went unnoticed by the health authorities either because they were not diagnosed or not reported to National TB Programs (NTP) (23). The notification rate increased from 2013-2015 according to the recent report by WHO due to the increase in notification rate (34%) by India. In 2015, 6.1 million new TB cases were notified out of 10.4 million incident cases (12). Even though TB diagnosis still depends on both the smear microscopy and sputum culture for identification in many countries, modern techniques are rapidly shifting the diagnostic landscape of tuberculosis.

The diagnosis of TB disease generally is based on a combination of clinical symptoms, chest X-ray examination and laboratory tests. In a TB dedicated laboratory clinical setting, diagnosis of active tuberculosis involves sputum smear microscopy, identification of the TB bacillus using culture techniques, phenotypic drug susceptibility tests and molecular tests. If infection is suspected then tuberculin skin test (TST), the interferon gamma release assays (IGRA) and acid fast staining of the sputum smear are conducted. (21).

1.6.1 Chest radiographs

Chest radiographs are important in the diagnosing of active tuberculosis with prominent features including upper lobe consolidation, cavitation and pleural effusions. Additional tests like computed tomography (CT) scanning are needed along with the other tests for the diagnosis of progressive primary tuberculosis or post primary tuberculosis which occurs at the first or second year after an initial infection (24,25).

Following chest radiographs or the initial diagnosis by the clinical symptoms, the common tests used as initial screening tests to differentiate the TB infection from disease are the tuberculin skin test, interferon-gamma release assays (IGRA) and Acid Fast Bacilli (AFB) smear staining (26).

1.6.2 Tuberculin Skin test (TST) or Mantoux test

TST is a preliminary test to identify people who are suspected to have TB infection. A sterile concoction of antigens from seven strains of *M. tuberculosis* called purified protein derivative (PPD) or tuberculin is injected intradermally to test the sensitivity of the skin to the PPD after 48 and 72 hours and the test is interpreted based on the size of induration. A positive reaction should be considered with caution as it does not exclude previous BCG vaccination, Mycobacteria other than tuberculosis (MOTT) infection and latent TB. Also due to the increased frequency of false negatives, a negative reaction does not exclude active or latent TB infection and it warrants extra tests.

1.6.3 Interferon- Gamma release assay (IGRA)

The IGRA test is used in diagnosing TB infections and has been reported to be more sensitive than TST for both active and latent pulmonary TB infections. It quantifies the T-cell response especially the interferon-gamma (IFNɣ) by either directly measuring the concentration of IFNɣ in the serum (Commercial kit: Quantiferon TB Gold) or by counting the number of T-cells that releases IFNɣ (Commercial kit: TSPOT.TB)(26).

1.6.4 Sputum microscopy

Sputum smear microscopy is preceded by acid-fast staining (Ziehl-Neelsen or Auramine fluorochrome stains are currently used) of clinical material and is the most important test for the diagnosis and screening of active tuberculosis in poor and low-income countries (27). WHO describes the presence of at least one acid fast bacillus (AFB) in at least one sputum sample as a smear positive pulmonary TB case. The guidance for interpreting smear results by WHO (a similar but slightly different classification is followed by Centers for Disease Control and Prevention in the USA) is given in table 1-2. It is widely used and effective in countries where TB is endemic and the sensitivity of microscopy has been reported to be higher than 80% (28,29) but a recent report by WHO states that it is around 20% (30). Sensitivity of the test decreases in HIV positive patients than in HIV negative cases and thus been associated with poor prognosis including death (31,32).

Table 1-2 Acid Fast Bacilli (AFB) Smear classification

(Adapted from WHO TB microscopy handbook)

Number of AFB seen per stated field (100x objective)	Smear Interpretation (Grade)
0 AFB in 100 fields	Negative
1-9 AFB in 100 fields	Record exact number of bacilli
10-99 AFB in 100 fields	1+
1-10 AFB / field, check 50 fields	2+
>10 AFB / field, check 20 fields	3+

1.6.5 Sputum culture

Culturing of mycobacteria follows the staining protocols either in a solid or a liquid medium after decontamination of the specimen. The recommended solid media by the European Respiratory Society, WHO, the American Thoracic Society and the International Union Against Tuberculosis (IUAT) are Lowenstein- Jensen (LJ) and Middlebrook 7H10 and 7H11 (MB) media (see fig 1-7 and 1-8). WHO has offered guidelines for the use of liquid media like *Mycobacterium* Growth Indicator Tube (MGIT) since they offer a more sensitive and quicker turnaround time than solid media (33,34).

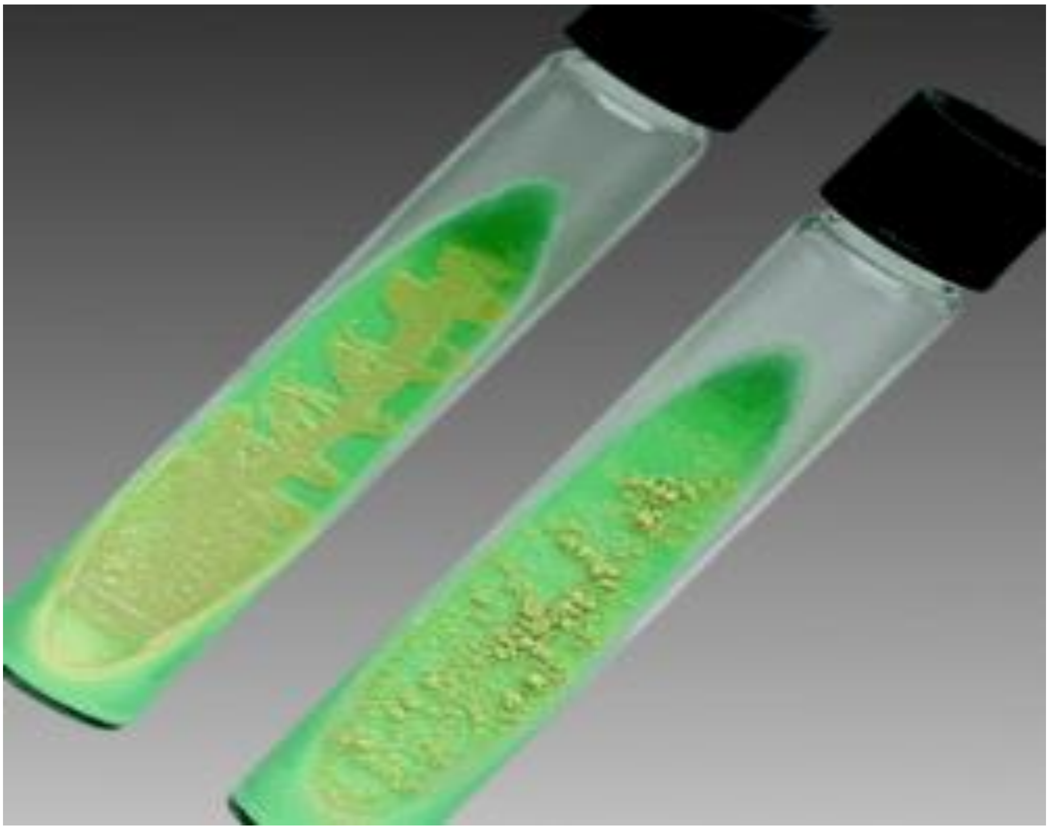


Figure 1-7 LJ medium showing dry, rough, raised creamy white or yellow colour colonies, characteristic of *M. tuberculosis*

(Taken from https://microbeonline.com/preparation-uses-lowenstein-jensen-lj-medium/)

Figure 1-8 Bactec MGIT system for detection of *M. tuber*culosis drug resistance

(Picture adapted from www.bd.com)

1.6.6 Molecular identification test and drug susceptibility tests

As drug resistance is a major problem in controlling tuberculosis especially with the rise in multidrug resistant (MDR) and extensively drug resistant (XDR) TB, drug susceptibility tests are carried out before starting treatment for tuberculosis. These tests are either phenotypic tests involving culture techniques or by adopting molecular tests. Phenotypic methods are based on testing sensitivity to first line anti-TB drugs and the most common being rifampicin, isoniazid, ethambutol and pyrazinamide. As these tests are time-consuming and may be ineffective in detecting low-level drug resistance, molecular tests are widely used. Commercial kits are available on the market based on the PCR amplification of specific genes including drug resistance genes and these molecular assays are grouped as either Nucleic Acid Amplification Tests (NAATs) or Line Probe Assays (LPA). The GeneXpert assay, a fully automated RT-PCR NAAT assay, is most suitable for patients with suspected pulmonary tuberculosis and for specific forms of extrapulmonary tuberculosis as in the case of meningitis in people with HIV and lymphadenitis, but not pleural, pericardial, or abdominal tuberculosis (34) and also determine the resistance to one of the most common TB drugs, Rifampicin (34). In countries like Tanzania which is a TB endemic country, the GeneXpert assay has been cost effective in determining the rifampicin-resistant status in the patient population (35). A second generation cartridge, the Xpert MTB/RIF Ultra (Ultra) is due to be launched on world TB day in 2017 by Cepheid and Rutgers and assessed in the TB centres globally using the current GeneXpert instruments (12).

1.6.6.1 Line Probe assays (LPA)

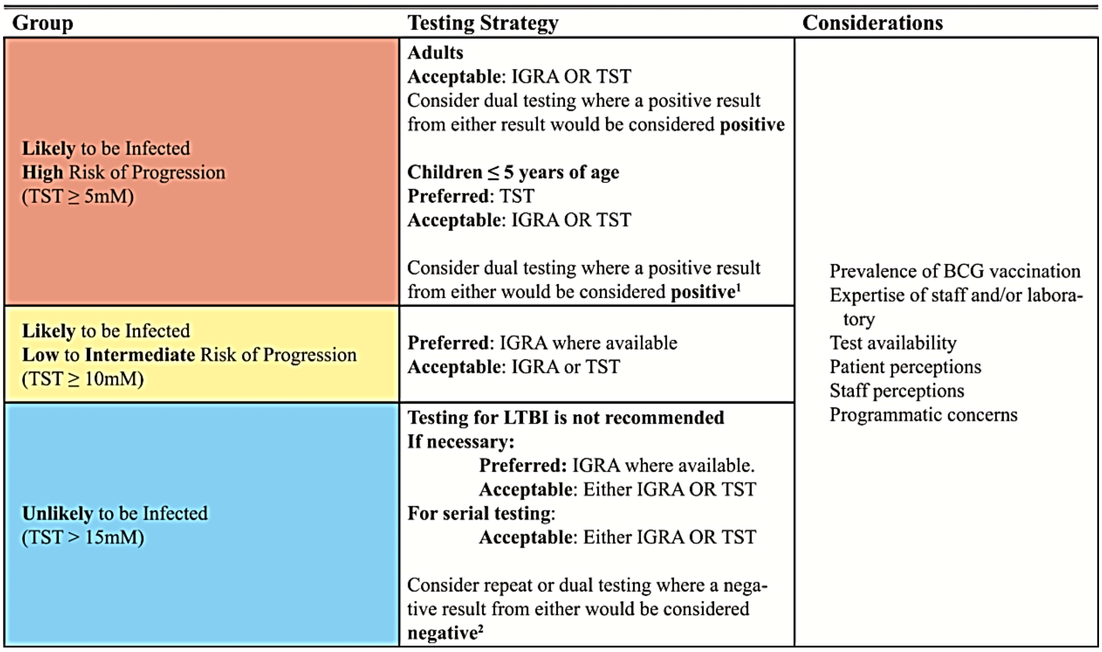
Line Probe Assays are based on the PCR amplification of specific fragments of the *M. tuberculosis* genome followed by hybridisation of PCR products to the oligonucleotide probes immobilised on membranes. The commercial kits that are available in the current market are INNO-LiPa Rif.Tb (Innogenetics, Belgium), GenoType MTBDR/ MTBDRplus (Hain Lifescience, Germany) and Genotype MTBDRsl (Hain Lifescience, Germany). GenoType MDRTBplus is the only kit that could detect isoniazid resistance and GenoType MTBDRsl detects resistance to fluoroquinolones and ethambutol , the second-line drugs that are used in MDR-TB and XDR-TB cases (36–40). There are several next generation technologies including GeneXpert cartridges and Hain MTBDRsl assays or rapid next generation whole genome sequencing methods, are in the advanced stages of development (41,42).

1.6.7 Diagnosis of Latent tuberculosis

Diagnosis of LTBI depends upon two factors: a. Likelihood of infection with *M. tuberculosis* and b. Likelihood of progression to TB disease (43). The treatment strategy for LTBI is shown below in table 1.3.

Table 1-3 Summary of recommendations for testing for latent tuberculosis infection (LTBI)

(taken from 43).



Group	Testing Strategy	Considerations
Likely to be Infected **High** Risk of Progression (TST ≥ 5mM)	**Adults** **Acceptable**: IGRA OR TST Consider dual testing where a positive result from either result would be considered **positive** **Children ≤ 5 years of age** **Preferred**: TST **Acceptable**: IGRA OR TST Consider dual testing where a positive result from either would be considered **positive**[1]	Prevalence of BCG vaccination Expertise of staff and/or laboratory Test availability Patient perceptions Staff perceptions Programmatic concerns
Likely to be Infected **Low** to **Intermediate** Risk of Progression (TST ≥ 10mM)	**Preferred**: IGRA where available **Acceptable**: IGRA or TST	
Unlikely to be Infected (TST > 15mM)	**Testing for LTBI is not recommended** **If necessary:** **Preferred:** IGRA where available. **Acceptable**: Either IGRA OR TST **For serial testing**: **Acceptable**: Either IGRA OR TST Consider repeat or dual testing where a negative result from either would be considered **negative**[2]	

Though conventional technologies still play a major role in the diagnosis of TB, the recent availability of new rapid tests has the potential to improve TB care. For example, the introduction of GeneXpert MTB/RIF assay using the GeneXpert instruments has enabled the rapid detection of tuberculosis in hours, in countries participating in the STOP TB campaign. In addition this system, by detecting rifampicin resistance more rapidly, has helped in the treatment of MDR-TB cases much more efficiently and also has been recommended for HIV-coinfected cases where smear microscopy tests could be less sensitive as described earlier (31,44). In the year 2015 alone, 6.2 million test cartridges were procured in 122 of the 145 countries eligible to purchase them at concessional prices showing a shift in the diagnosis landscape of tuberculosis (12). There is still a requirement for newer technologies which are much more affordable than the currently available strategies (45).

The advancement in new diagnostics and drugs that has been progressing over the last decade would help in combatting tuberculosis by improving the diagnosis of the disease and also early detection of drug resistant tuberculosis (37,46). As per the review by Drobniewski et al., it is important to identify TB cases including the drug resistant strains, early and accurately (37) and to arrest TB transmission by providing appropriate therapy and curing patients.

1.7 TB treatment

A minimum of six months’ treatment is recommended which is divided into two phases, the initial phase and the continuation phase (44). The standard treatment regimen according to WHO (47) for both pulmonary TB including HIV-coinfected individuals and for extra-pulmonary TB cases is as follows:

a. Initial phase: Daily medication or therapy five days/week under directly observed therapy (DOTS) with isoniazid (INH), rifampicin (RIF), pyrazinamide (PZA) and ethambutol (EMB) for 2 months. A sputum smear test is taken and the next phase is only started if the test is negative.

b. Continuation phase: Daily medication or therapy five days/ week under DOT with isoniazid and rifampicin for the next four months.

According to the review by Fonseca et al., (48) multidrug resistant (MDR) and extensively drug resistant (XDR) TB strains pose a significant threat to the control of tuberculosis as they develop resistance to first line drugs mentioned above and work still needs to be done to understand the mechanism of resistance. If there is resistance to two of the first line drugs, RIF and INH, then these strains are termed as MDR- TB strains. The treatment course for MDR-TB lasts for more than 2 years with second line TB drugs like fluoroquinolones and injectable aminoglycosides like streptomycin. Anti-TB drug resistance occurs due to the chromosomal mutations at a rate of 10^{-3} to 10^{-9} in genes encoding drug targets (49,50). If the treatment is not completed or if the proper regimen is not followed then it results in the poor treatment of MDR-TB patients thus amplifying the resistance of the first line drugs and leading to the development of XDR-TB (34).

1.7.1 Rifampicin (RIF)

Rifampicin inhibits RNA transcription and thereby protein synthesis by binding to the *rpoB* encoded subunit of RNA polymerase. Mono resistance to RIF is a rare occurrence as nearly 85-90% of RIF resistant strains are also resistant to INH. RIF resistance therefore acts as a surrogate marker for MDR-TB (51–53). A high percentage of mutations (90-95%) occurs in the 81bp rifampicin resistance-determining region (RRDR) of the *rpoB* gene between codons 507-533 and a lower level of resistance (5-10%) is seen in the N-terminal or cluster II region of rpoB / other genes at codon V146 (51,54). It has been suggested that higher frequency of mutation is seen in some strains at certain positions of the RRDR, especially at the codon 531 where serine is replaced by leucine, due to the higher relative strain fitness than other strains (55).

1.7.2 Isoniazid (INH)

INH a bactericidal drug which targets the NADH- specific enoyl acyl carrier protein (ACP) reductase (*inhA*) and prevents the production of mycolic acid which leads to the inhibition of dividing mycobacterial cells (51,56). INH resistance is commonly noticed due to mutations within the *katG* gene at codon 315 or low level resistance mutations are seen in the *inhA* regulatory region (51,56,57). However, oxidative stress does not seem to impair the mutation rates in Isoniazid resistant strains (58).

1.7.3 Pyrazinamide (PZA)

PZA, activated by pyrazinamidase (*pncA*), is effective against *M. tuberculosis* by inactivating the fatty acid synthesis pathway. In addition, it also prevents protein translation and the ribosome sparing process of translation by binding to ribosomal protein S1 (*rpsA*) in *M. tuberculosis*. PZA resistance is seen in strains containing mutations in *pncA* or *rpsA* (51,59).

1.7.4 Ethambutol (EMB)

EMB prevents the formation of cell wall assembly by inhibiting the synthesis and polymerisation of cell wall arabinan leading to the accumulation of free mycolic acid. Resistance mutations are commonly seen in three *emb* genes at codons 306, 406 and 497(60).

1.7.5 Second line drugs

Some of the second-line drugs used in TB treatment, which include Streptomycin (SM), Kanamycin (KAN) and Amikacin (AMI) inhibit protein synthesis and are bactericidal drugs. Fluoroquinolones are also widely used as anti-TB drugs especially Oflaxacin (OFX), Levofloxacin (LFX) as bacteriostatics and Gatifloxacin (GFX) and Moxifloxacin (MFX) as bactericidals by inhibiting DNA gyrase leading to cell apoptosis by preventing

DNA replication and repair as shown by the Rapid Evaluation of Moxifloxacin in Tuberculosis (REMox TB) phase 3 study trial by Gillespie et al (61,62).

1.7.6 Vaccination

The BCG (Bacillus Calmette-Guerin) vaccine, designed to prevent tuberculosis is derived from continuous passaging of an *M. bovis* strain and has been in the public health domain for the last eighty years (63,64) This vaccine induces a strong T-cell response by activating the Th1 cells but it offers only limited protection in adults against pulmonary tuberculosis as the immune response is not prolonged enough to prevent *M. tuberculosis* infection developing in older individuals(65,66).(64), but is effective in infant TB meningitis (67). In the last decade there has been continuous research to develop an effective vaccine that could either complement the BCG vaccine or act independently to prevent TB infection and provide a long- term protective immunity. The Modified- Vaccinia – Ankara (MVA) 85A vaccine that was trialled in children, was found to be ineffective in children in offering protection against TB (68).

1.8 Molecular markers in TB epidemiology

Various molecular tools have been employed to study the epidemiology of tuberculosis based on the availability of the instruments and the ability to genotype clinical isolates. The molecular tools are broadly divided into non-PCR based tools - Restriction Fragment Length Polymorphism (RFLP) and Polymorphic GC-rich Repetitive Sequences RFLP (PGRS-RFLP) and PCR based tools (Spoligotyping, Mycobacterial Interspersed Repetitive Unit –

Variable Number Tandem Repeat (MIRU-VNTR), Amplified Fragment Length Polymorphism (AFLP) and Whole Genome Sequencing (WGS) methods.

1.8.1 IS*6110* and IS*6110* RFLP

Bacterial transposons and Insertion Sequences (IS) are evolutionarily informative as they alter the genetic makeup of the host organisms during transposition (69). Insertion sequences are mobile genetic elements that code for transposases evolved to move the IS around the genome. A study in 1991 (70) showed that the *M. bovis* BCG insertion element IS*987* is virtually identical to the previously described IS elements IS*986* (71) and IS*6110* from *M. tuberculosis.* The only biologically significant difference is the presence, in IS*987,* of ORFa (Open Reading Frame) in one single ORF, whereas IS*986* and IS*6110* contain ORFa, composed of two different ORFs (70,72) (see figure 1.9).

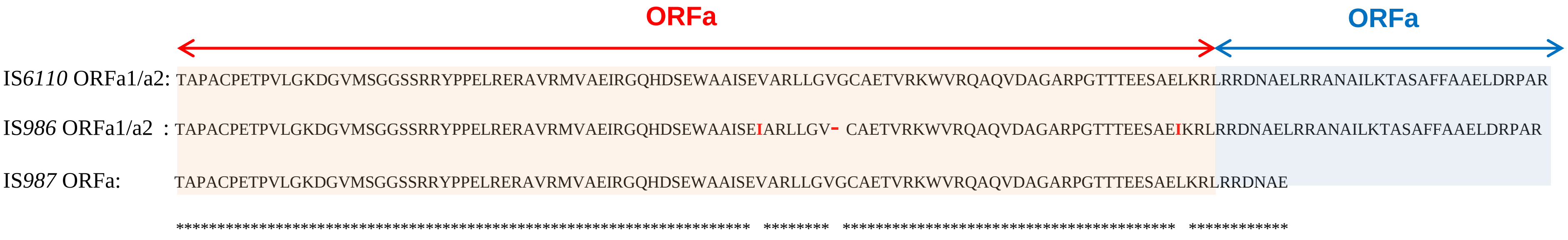


Figure 1-9 Translated sequence alignment of ORFa in IS*6110* and IS*986* both from *M. tuberculosis* and IS*987* from *M. bovis* BCG

(Asterisks denotes the matching amino acid residues in all three sequences) (Adapted from (70)).

Tracing the movement of these elements in the host genomes is productive in the identification of different strains of microorganisms. IS can transpose within the genomes thereby altering the position and even function of the adjacent genes, for example insertion of an insertion element (IE), IS*200-like* element, into the genome of *Yersinia pestis* inactivates the *invasin* gene which enables the bacteria to invade the host organism (73). IS*6110* insertions have been found in some of the essential genes in drug resistant strains (74). These changes contribute to the evolution of microorganisms introducing genomic plasticity facilitating survival in different environmental conditions (75).

IS elements have been used as taxonomic markers over the years in certain bacteria like *Bordetella pertussis*, *Salmonella typhi* etc. (76). An IS-like element IS*6110*, with inverted (28bp with 3 mismatched bases) and direct (3bp)repeats at its ends, was identified as a repeated sequence from an *M. tuberculosis* cosmid library constructed in pHC79 (77,78) by screening the library with labelled *M. tuberculosis* total DNA. This element is a 1.36kb insertion sequence found only in MTBC members and belongs to the enterobacterial Insertion Sequence 3 (IS*3*) family of insertion sequences. It contains an ORF encoding transposase of 1037bp length and has 28bp imperfect inverted repeats at both the ends of the sequence (78) (fig 1-10).

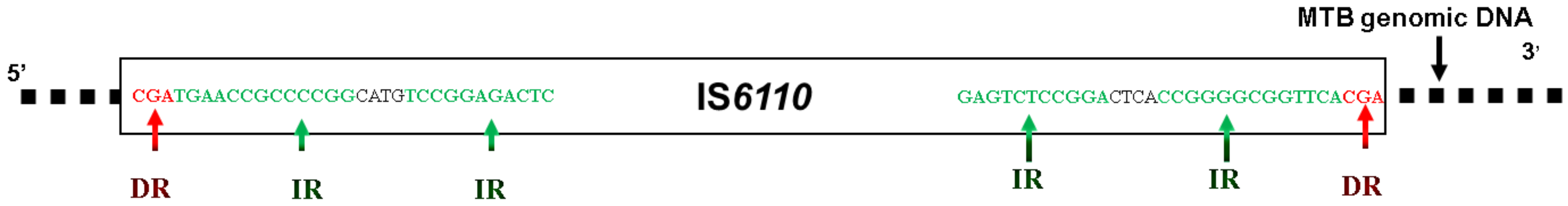


Figure 1-10 Pictorial representation of IS*6110* sequence of *M. tuberculosis* genome showing Direct (DR in red colour) and inverted repeat (IR in green colour) sequences

The first instance of the use of IS*6110* as a probe for the identification of MTBC was reported in 1990 (78). MTBC from uncultured specimens, were identified using primers designed from the IS*6110* sequence in a PCR reaction and results were successfully correlated with other identification procedures (78). This element is still being used as a tool for DNA fingerprinting to the present day as it is conserved in the MTBC (79,80). IS*6110* copy numbers vary from strain to strain; one in *M. bovis* and >25 in *M. tuberculosis* (81). Also, three to four nucleotides of the genomic sequence are repeated at the extremities of some IS*6110* copies or the site of insertion suggesting a role of transposition of these elements in MTBC genomes (82). An oligonucleotide ligation assay (OLA) using fluorescently labelled IS*6110* oligonucleotides was employed to detect MTBC (83) Since then the location of the IS*6110* element has been used a powerful tool for the rapid fingerprinting of isolates of MTBC (84–86), including isolates of *M. bovis* (87,88). As there is variation in the copy numbers of IS*6110* in different strains and because IS*6110* does not induce *in vivo* any major genomic rearrangements for approximately five to eight years (90,91). IS*6110* Restriction Fragment Length Polymorphism (RFLP) was the ‘gold standard’ typing method in strains with more than five copies before being replaced by the MIRU-VNTR technique (92–94) (see section 1.8.4). In a study reported by Warren et al in 2002, it was found that the rate of transposition occurs at a rate of 18.6% for every 6.5 years i.e. 2.9% per year, thereby altering the genotype of the organism (91) It has been suggested that there might be differences between different lineages, one change every 3 years in Euro-

American lineage and it might take ten times longer to observe one change in East Asian lineage (95).

Also, comparison of the number and length of fragments generated due to IS*6110* insertion in isolates proved to be discriminatory (96). Sun et al., in 2004 showed that spoligotyping and Variable Number Tandem Repeats (VNTR) (see section 1.8.4) analysis can further support the resolution of IS*6110*-defined low copy number isolates. With these advantages put aside, IS*6110* RFLP is however considered a laborious procedure involving steps from culturing the mycobacteria, extraction of DNA, restriction enzyme digestion, Southern Blotting and IS*6110* probing.

1.8.1.1 Role of IS*6110* in the biology of *M. tuberculosis*

The advantages that transposition of this element brings to the host either via gene disruption, gene excision or by enhanced gene expression suggests that they offer the selective advantage to strain fitness and may influence the genomic plasticity of these organisms. It can up-regulate downstream genes through an outward-directed promoter in its 3' end, thus adding to the significance of this element. The ability to activate genes during infection suggests that IS*6110* might have the potential to influence growth characteristics of different strains, and indicate another mechanism by which IS*6110* can impact *M. tuberculosis* evolution (97). Gene disruption or excision leads to loss of gene expression and enhanced gene expression leads to over - expression of genes as seen in the Proline Glutamate-polymorphic GC rich Sequences (PE-PGRS gene -Rv1468c). This effect of supplying an outward-directed promoter-like sequence is also seen in *ctp*D gene (a cation transporting ATPase) of a Beijing genotype useful for *ctp*D transcription (97,98). It is reported that the *M. tuberculosis ctp*D orthologue resists cellular redox stress by controlling the effects of cobalt and nickel ions on the bacterial cell (99). Several studies have shown that IS*6110* insertion does not affect the replication machinery or the growth patterns but suggest that it definitely helps the organisms to adapt to the host and the environment. Transposition of IS elements varies from element to element within the genome (100) as an element situated within a transcriptionally active site or downstream from an external promoter will be more mobile than the ones located in the non–coding regions.

Transposition of IS*6110* occurs at random, independent of IS preferential loci or hot-spots. However, sites of integration were not thought to be entirely random as different ‘hot-spot’ insertion sites have been identified previously that include the DR locus (70), the *ipl* locus (Fang and Forbes, 1997; Fang et al., 1999), dnaA- dnaN intergenic region (103), the phospholipase region (104) and the region between Rv1754c and Rv1762c (105). The property of transposition rendering a distinct signature to the genomes acts as a suitable genetic marker for studying the divergence of MTBC from its common ancestor. Mapping this element in the genomes of MTBC found that it does not favour particular nucleotide sequences (82,106) for its insertion. Insertion at the DR locus is unique as it is found in a majority of the *M. tuberculosis* complex strains especially in *M. bovis* in that it is its only site of insertion. It is also considered to be the primordial insertion site in the common ancestor before the two species diverged with further copies appearing due to the transposition of the element outside the DR region (70). In low copy strains, lack of mobility of the IS element due to a low transcriptional activity at the site of insertion could play a role in the lack of IS*6110* RFLP pattern diversity (107). It has been shown that insertion favours regions with low GC content as observed in the case of insertion into the PE/PPE gene family where PE (75% GC) is less favoured than PPE (64%) which was also complimented by another study showing a mini-transposon Tn5370 favouring lower GC content genomic regions in *M. tuberculosis* (98,106,108). When there are five or less copies of IS*6110*, this DNA fingerprinting technique is however not very reliable. So the use of polymorphic GC rich repetitive sequences (PGRSs) were highlighted along with DR spoligotyping and 16S rRNA sequences in tandem to identify strains containing few copies of IS*6110*, as in *M. bovis* or strains with no copy of IS*6110* in their genomes (107,109) Dale et al. (107) suggests that there could be a low mobility rate of transposition in strains with low copy numbers.

1.8.2 PGRS- RFLP

A specific 3.4kb fragment of the PGRS cloned in plasmid pTBN12 was used as a probe for differentiation of unrelated strains of TB and identification of related strains using identical banding patterns (110). Interpretation of the banding patterns is more difficult as it does not contain as many variations as other markers like IS*6110* and also requires a large amount of DNA as well as being laborious and time consuming and so this technique is no longer used (111).

1.8.3 Spoligotyping

All the members of the MTBC have the Direct Repeat (DR) locus in their genomes. The DR locus is a member of the CRISPR family (Clustered regularly interspaced short palindromic repeats) and is susceptible to polymorphic changes. This property was exploited for a genetic strain typing technique called spacer-oligo typing or spoligotyping (112,113). There are 43 unique non-repetitive spacer sequences (35-45bp in length) interspersed with direct repeat sequences (36bp in length) known as direct variable repeat (DVR) sequences. Due to the loss of single or consecutive DVR sequences caused either by homologous recombination between neighbouring or distant direct repeat sequences or by the loss of sequences caused by the excision of the IS*6110* element, this method is useful for discriminating between strains. As the DVR sequences are numbered 1-43 and are well conserved between strains, spoligotyping is quite a quick and useful method (see figure 1-11 below) to interpret into either octal or hexadecimal codes (102,114).

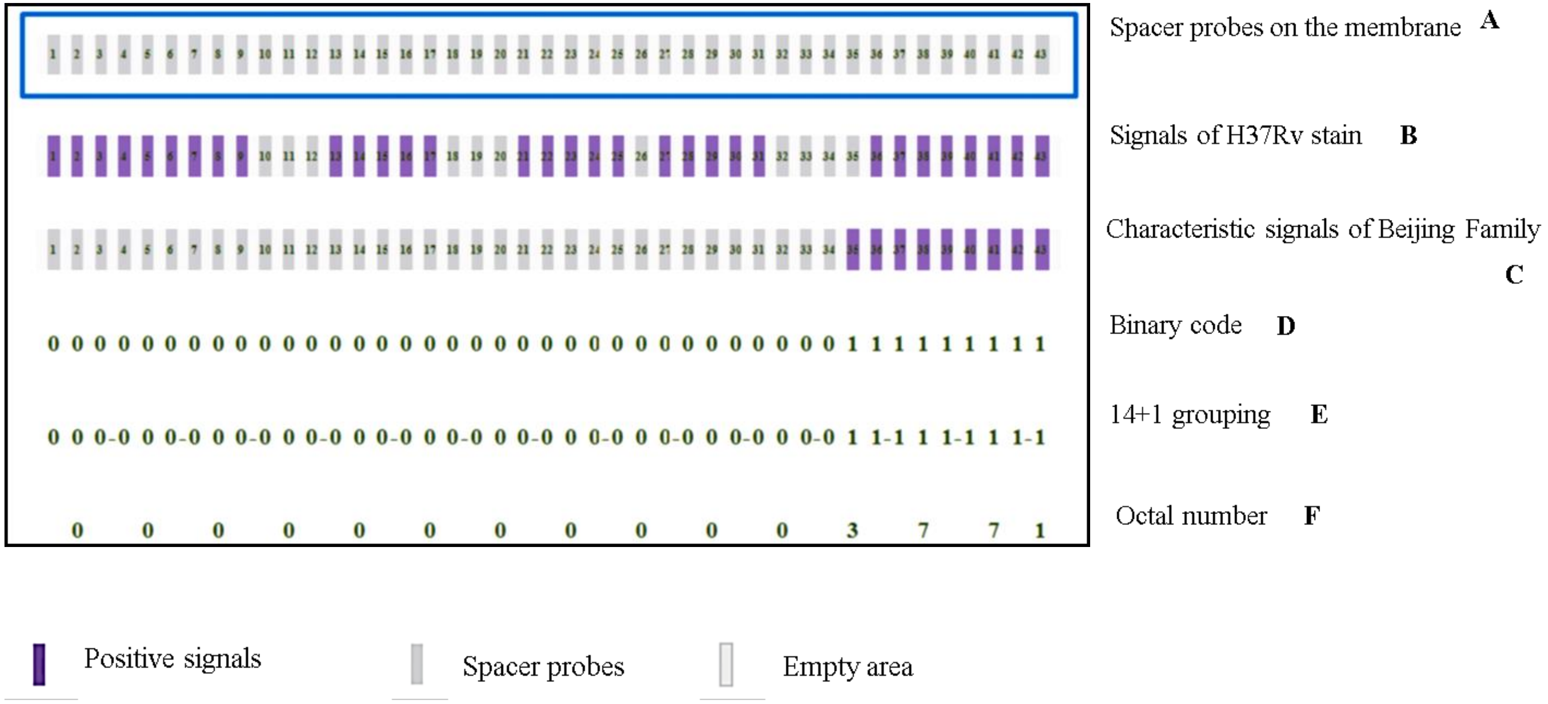
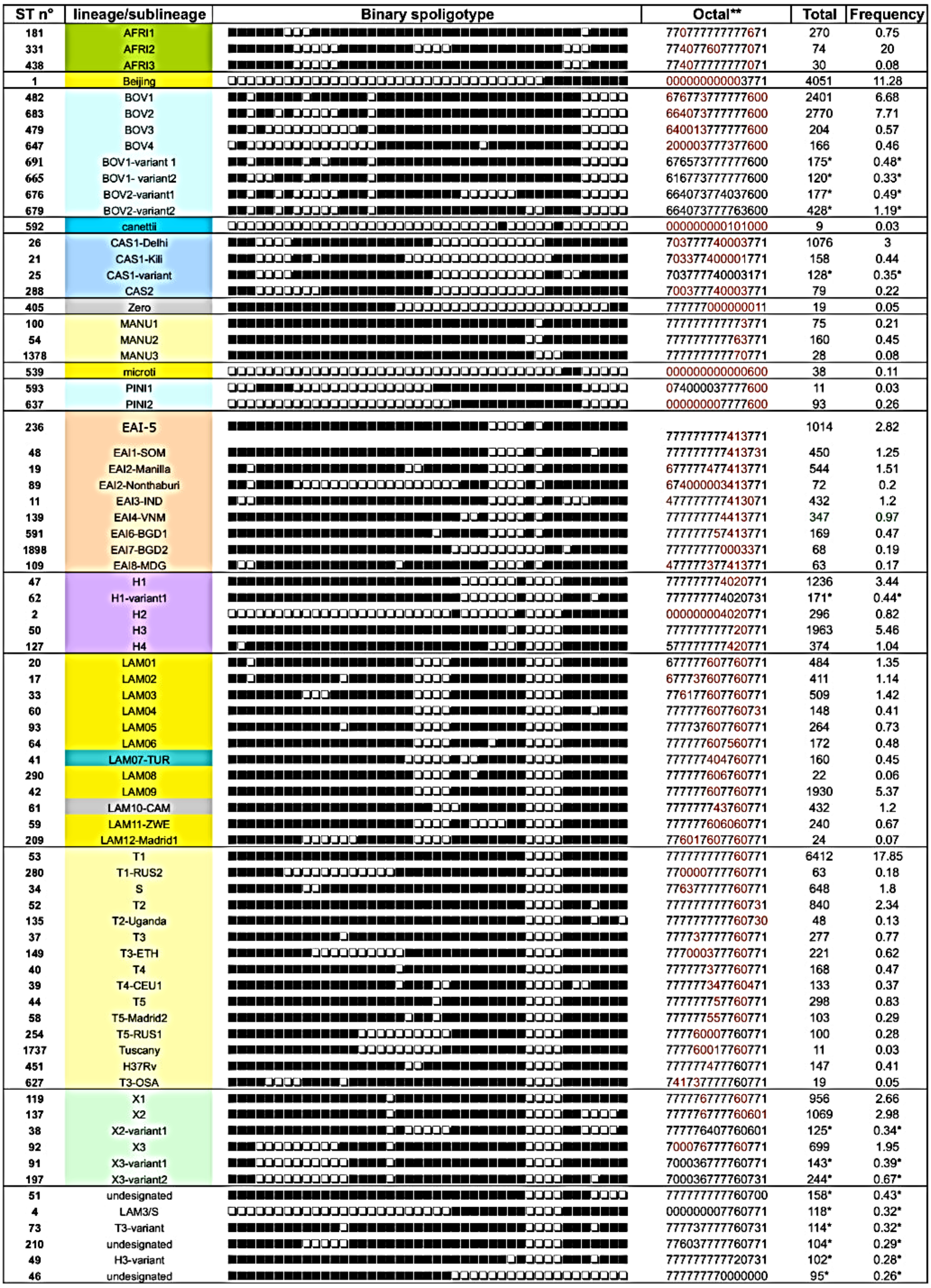


Figure 1-11. Principle of spoligotyping and the processing of signals

(A) *M. tuberculosis* genome with well-conserved 36-bp direct repeats (DRs) which are interspersed by 35-43 bp of unique spacer sequences. Genetic diversity depends on the deletion of these spacer regions. The spacer regions are amplified by primers, and the presence of at least one spacer fragment shows a PCR positive reaction. On the membrane, 43 probes targeting each spacer are spotted, and a unique pattern of spoligotyping is visualized after hybridization with PCR product. (B) Signals of reference strain H37Rv. (C) A typical signal pattern of Beijing family *M. tuberculosis* strain. (D) To analyze signal patterns, the signals are converted to binary code of ‘on (1) and off (0)’. (E) The 43-digit binary code is converted to a 15-digit octal (i.e., base 8, having the digits 0-7) designation by a two-step process. First, the 43-digit binary code is divided into 14 sets of three digits (spacers 1 through 42) plus one additional digit (spacer 43). (F) Each 3-digit binary set is converted to its octal equivalent, with the final additional digit remaining as 1 or 0. The translation of binary numbers to octal numbers is done as follows: 000 = 0; 001 = 1; 010 = 2; 011 = 3; 100 = 4; 101 = 5; 110 = 6; and 111 = 7 (Taken from (115))..

Spoligotyping is a simple PCR based reverse hybridisation blotting technique where the DR locus is first amplified followed by the hybridisation with 43 synthetic spacer oligonucleotide probes covalently bound to the nylon membranes is visualised using a chemiluminescence system (112,115) for the presence of a 'dark' band (presence of a spacer) or 'no' band (absence of a spacer). There are at least nine spoligotype-defined families with specific hybridisation signatures (Haarlem, Beijing, LAM, CAS, EAI, S,T, X and AFRI families as seen in figure 1.12 which can be further divided into 36 subfamilies of the MTBC (113,116). This technique has a lower sensitivity than IS*6110* RFLP but is useful in delineating strains with fewer than five IS*6110* copies (117,118) and is the gold standard typing method for *M. bovis* strains lacking spacers 39-43.

Currently there is a database which can be accessed online (http://www.pasteur-guadeloupe.fr.8081) containing both octal and binary (as seen above in figure 1.12) spoligotype descriptions for strains of the MTBC species isolated globally known as SpolDB4 (113) (see figure 1-12). It contains 1939 shared-types (STs) representing a total of 39,295 clinical isolates originating from 122 countries. Due to the low resolution of the marker when used alone, a publicly available database known as SITVITWEB that utilises SpolDB4 data along with MIRU-VNTRs (see section 1.8.4 below) is in place for high-resolution epidemiological studies (119).



ST n°	lineage/sublineage	Binary spoligotype	Octal**	Total	Frequency
181	AFRI1		770777777777671	270	0.75
331	AFRI2		774077607777071	74	20
438	AFRI3		774077777777071	30	0.08
1	Beijing		000000000003771	4051	11.28
482	BOV1		676773777777600	2401	6.68
683	BOV2		664073777777600	2770	7.71
479	BOV3		640013777777600	204	0.57
647	BOV4		200003777377600	166	0.46
691	BOV1-variant 1		676573777777600	175*	0.48*
665	BOV1- variant2		616773777777600	120*	0.33*
676	BOV2-variant1		664073774037600	177*	0.49*
679	BOV2-variant2		664073777763600	428*	1.19*
592	canettii		000000000101000	9	0.03
26	CAS1-Delhi		703777740003771	1076	3
21	CAS1-Kili		703377400001771	158	0.44
25	CAS1-variant		703777740003171	128*	0.35*
288	CAS2		700377740003771	79	0.22
405	Zero		777777000000011	19	0.05
100	MANU1		777777777773771	75	0.21
54	MANU2		777777777763771	160	0.45
1378	MANU3		777777777770771	28	0.08
539	microti		000000000000600	38	0.11
593	PINI1		074000037777600	11	0.03
637	PINI2		000000007777600	93	0.26
236	EAI-5		777777777413771	1014	2.82
48	EAI1-SOM		777777777413731	450	1.25
19	EAI2-Manilla		677777477413771	544	1.51
89	EAI2-Nonthaburi		674000003413771	72	0.2
11	EAI3-IND		477777777413071	432	1.2
139	EAI4-VNM		777777774413771	347	0.97
591	EAI6-BGD1		777777757413771	169	0.47
1898	EAI7-BGD2		777777770003371	68	0.19
109	EAI8-MDG		477777377413771	63	0.17
47	H1		777777774020771	1236	3.44
62	H1-variant1		777777774020731	171*	0.44*
2	H2		000000004020771	296	0.82
50	H3		777777777720771	1963	5.46
127	H4		577777777420771	374	1.04
20	LAM01		677777607760771	484	1.35
17	LAM02		677737607760771	411	1.14
33	LAM03		776177607760771	509	1.42
60	LAM04		777777607760731	148	0.41
93	LAM05		777737607760771	264	0.73
64	LAM06		777777607560771	172	0.48
41	LAM07-TUR		777777404760771	160	0.45
290	LAM08		777777606760771	22	0.06
42	LAM09		777777607760771	1930	5.37
61	LAM10-CAM		777777743760771	432	1.2
59	LAM11-ZWE		777777606060771	240	0.67
209	LAM12-Madrid1		776017607760771	24	0.07
53	T1		777777777760771	6412	17.85
280	T1-RUS2		770000777760771	63	0.18
34	S		776377777760771	648	1.8
52	T2		777777777760731	840	2.34
135	T2-Uganda		777777777760730	48	0.13
37	T3		777737777760771	277	0.77
149	T3-ETH		777000377760771	221	0.62
40	T4		777777377760771	168	0.47
39	T4-CEU1		777777347760471	133	0.37
44	T5		777777757760771	298	0.83
58	T5-Madrid2		777777557760771	103	0.29
254	T5-RUS1		777760007760771	100	0.28
1737	Tuscany		777760017760771	11	0.03
451	H37Rv		777777477760771	147	0.41
627	T3-OSA		741737777760771	19	0.05
119	X1		777776777760771	956	2.66
137	X2		777776777760601	1069	2.98
38	X2-variant1		777776407760601	125*	0.34*
92	X3		700076777760771	699	1.95
91	X3-variant1		700036777760771	143*	0.39*
197	X3-variant2		700036777760731	244*	0.67*
51	undesignated		777777777760700	158*	0.43*
4	LAM3/S		000000007760771	118*	0.32*
73	T3-variant		777737777760731	114*	0.32*
210	undesignated		776037777760771	104*	0.29*
49	H3-variant		777777777720731	102*	0.28*
46	undesignated		777777770000000	95*	0.26*

Figure 1-12 Pictorial representation of SpolDB4 classification analysis of different TB lineages

First column ST n°: Shared-type (ST) number of prototype pattern for the linage/sub lineage. Second column: lineage/sub lineage name. Third column: Binary spoligo display with black-white squares for respectively hybridizing-non-hybridizing spacers. Fourth column: Octal code (in red: defining octal rule). Fifth column: total absolute number of isolates of the subclass when variant ST spoligos are included (using SpolNet). Sixth column: same but expressed as percentage of total clustered isolates. * Total number and Frequency for these types are already included in their mother clade if known. Undesignated types are counted within the T1-ill-defined lineage. ** in red: octal rule defining the genotype (Taken from (113)).

1.8.4 MIRU-VNTR

Mycobacterial interspersed repetitive units (MIRUs) in *M. tuberculosis* are grouped under minisatellites which are short highly repetitive DNA sequences and are 40-100 bp long. They are suggested to have a role in chromosome structure and rearrangement, tandem duplications, differential translation, transcriptional termination and antigenic variation (120). MIRUs are highly polymorphic at tandem repeat loci, and are useful in DNA typing studies in both prokaryotes and eukaryotes. They can also be referred to as variable number tandem repeats (VNTRs) and exact tandem repeats (ETRs). ETRs are 53-79bp long tandem repeats and another repeated region called Major Polymorphic Tandem Repeats (MPTRs) were first described in 1998. Some MIRU loci are positioned in the polycistronic operons and variability in copy number can affect the expression of the flanking genes (121).

Mycobacterial Interspersed Repetitive Units (MIRU) first described by Supply et al., in 2000 and the Exact Tandem Repeats (ETR) both comprise MIRU-VNTR, a typing technique useful in epidemiology, population genetics and phylogenetic studies (122). The technique involves the PCR amplification of the entire tandem repeat loci using primers in the DNA sequences flanking the repeats. The number of repeats or alleles is calculated from the amplicon sized by electrophoresis as the length of the repeats and the position of the primers are known. Twelve MIRU loci (2, 4, 10, 16, 20, 23, 24, 26, 27, 31, 39, 40) and five ETR loci have been used previously for this genetic fingerprinting technique. This method was widely adopted in the UK. There is some overlap of the nomenclature of these repeat schemes, for example MIRU 4 and MIRU 31 are the same loci as ETR-D and ETR-E respectively.

Supply et al., Ii 2006 proposed a 24 loci typing technique comprising of 12 MIRU loci, 3 ETR loci, six 'Mtubs' and three 'QUB'loci. This was the gold standard typing method used in the UK until recently but now WGS has become the method of choice in the UK. A fifteen loci subset system with the highest evolutionary rates was proposed as the best system for first-line analysis of outbreak samples and the 24 loci subset for the phylogenetic studies which need higher resolution. The twelve loci subset, analysing the most variable regions, are also used for initial investigations along with data from other techniques like spoligotyping and *IS6110* fingerprinting (123).The following table 1-4 shows the widely used MIRU-VNTR schemes worldwide.

Table 1-4 Composition of the different MIRU-VNTR sets widely used

(Taken from http://www.miru-vntrplus.org/MIRU/miruinfo.faces)

Loci	Alias 1	Alias 2	24	15	12
154	MIRU02		X		X
424	Mtub04		X	X	
577	ETR-C		X	X	
580	MIRU04	ETR-D	X	X	X
802	MIRU40		X	X	X
960	MIRU10		X	X	X
1644	MIRU16		X	X	X
1955	Mtub21		X	X	
2059	MIRU20		X		X
2163b	QUB11b		X	X	
2165	ETR-A		X	X	
2347	Mtub29		X		
2401	Mtub30		X	X	
2461	ETR-B		X		
2531	MIRU23		X		X
2687	MIRU24		X		X
2996	MIRU26		X	X	X
3007	MIRU27	QUB5	X		X
3171	Mtub34		X		
3192	MIRU31	ETR-E	X	X	X
3690	Mtub39		X	X	
4052	QUB26		X	X	
4156	QUB4156		X	X	
4348	MIRU39		X		X

1.8.5 AFLP and FAFLP

1.8.5.1 AFLP

Amplified Fragment Length Polymorphism (AFLP) was designed as an identification and typing method developed in the nineties by the biotechnology company called Keygene (124,125). It is a selective restriction fragment amplification technique based on the ligation of adapters to genomic restriction fragments and PCR amplification with adapter specific primers. A small amount of genomic DNA is digested with two restriction enzymes and then ligated with double-stranded oligonucleotide adapters designed specifically with a point mutation so that the initial restriction is not restored after ligation. With the aid of stringent PCR conditions and adapter-specific primers at the 3’ ends, one of the three nucleotides is extended from the unknown restricted genome fragment. The primers initially were radioactive labelled for easy identification of the fragments in polyacrylamide gels which were later replaced by fluorescently labelled primers for visualisation in automated platforms (125–127).

Since then, AFLP has been used as a diagnostic tool in plants and animals in the field of genetic mapping, phylogenetic studies, microbial typing and for diagnostic purposes (124,128,129).

1.8.5.2 FAFLP

Fluorescent Amplified Fragment Length Polymorphism (FAFLP) exploits the use of fluorescently labelled primers and the resulting genomic restriction fragments are visualised using an automated platform. Since the restricted fragments with base substitutions are visualised using a platform such as the

ABI Genetic Analyser (Life Technologies, UK), this technique can give a snap shot of the insertion variation on a genomic level. It has been shown that a single fragment difference (± 1 bp) in the FAFLP profile signifies a new strain (130). FAFLP should, like all other techniques, be considered in tandem with epidemiological data if available. FAFLP has been successfully applied to several organisms and some of them have been shown below in the form of a table (see table 1-5). It has also been used in the identification of plant species with 98% accuracy (131).

Table 1-5 List of some common microorganisms studied using FAFLP analysis

Microorganism	**References**
Campylobacter jejuni	(132)
Escherichia coli K12 MG1655	(127)
Escherichia coli O157	(127,133)
Listeria monocytogenes	(134)
Mycobacterium tuberculosis	(130,135)
Neisseria gonorrhoeae	(136)
Neisseria meningitides	(137)
Streptococcus pyogenes serotype M1	(126)

Staphylococcus aureus	(138)
Bacillus thuringiensis	(139)
Candida tropicalis	(140)
Acanthamoeba Sp.	(141)
Salmonella enteritidis	(142)
Leptospira borgpetersenii serovar Arborea	(143)
Pseudomonas aeruginosa	(144)
Legionella pneumophila serogroup 1	(145)

1.8.5.3 FAFLP for TB

1.8.5.3.1 IS*6110* FAFLP

IS*6110* RFLP is a time consuming and low-throughput technique. To overcome these drawbacks and utilise the superior resolution property of IS*6110,* IS*6110* FAFLP was developed. FAFLP derived sequence data is congruent with the IS*6110* typing of *Mycobacterium tuberculosis* but it should be used along with other techniques like spoligotyping and MIRU-VNTR to supplement epidemiological data (Goulding, Stanley, et al. 2000; Thorne et al. 2007). In a study by Thorne et al, in 2007, it was noticed that the FAFLP results of 57 (97%) of the 59 clustered isolates were congruent with the RFLP results. Use of coloured fluorescent primers can differentiate between recent transmission of strains and epidemiologically unrelated but genetically related strains (147). IS*6110* FAFLP has been used recently to identify common fragments in MTBC that likely evolved from common ancestors and were thereby able to differentiate the strains phylogenetically into different TB lineages (135). This study also indicated that IS*6110* FAFLP contains a strong phylogenetic signal in modern TB lineages by assigning common fragments to their respective spoligotypes/Principal Genetic Groups (PGGs).

1.8.5.4 DNA Sequencing

From Sanger sequencing to Next generation sequencing technologies, there are different sequencing methodologies that can be used to generate data for molecular epidemiological studies of tuberculosis (148). Due to the high cost and low quality of the sequences both in the first 15-40bp and after 700bp of sequenced product, Sanger sequencing is being replaced by Next Generation Sequencing (NGS), also known as Whole Genome Sequencing

(WGS). Over the last ten years, platforms including, Illumina, Roche, ABI SOLID, ION PGM, Heliscope, Pacific Biosciences Single Molecule Real Time (SMRT) and the Nanopore MinION (149) have been used. WGS is likely to become the gold standard approach in the future but specialist software to analyse the data and a skilled workforce are still required to interpret the results.

1.9 Global TB lineages

New species of mycobacteria have evolved over the years by many adaptive changes to different niches (6). This has led to the reduced function or inactivation of certain genes. A classic example is *M. leprae* which has transformed to become an obligate intracellular pathogen by losing many genes involved in metabolic and respiratory pathways (150). Genes can also be acquired by some species like *M. abscessus* which is one of the organisms responsible for causing infection in those suffering from cystic fibrosis. This organism was thought to originally be a soil saprophyte, and has acquired essential genes to survive phagocytosis enabling intra-cellularisation in host organisms. Recently it has been shown that *M. abscessus* has acquired resistance to arsenic and mercury which are typically found in resistance plasmids of environmental organisms like *M. marinum* through horizontal gene flux (151).

All the members of the MTBC share identical 16S rRNA sequences and more than 99% similarity at the genomic level and yet are phenotypically different causing different pathologies in different host species. *M. canetti* is unique in this group in that it does not share homogeneity with the MTBC in some house-keeping genes, Insertion Sequence *1081* copy number, colony morphology and lipid content of the cell. Brosch et al., in 2002 showed that *M. canetti* might have diverged directly from the common ancestor of MTBC some 30,000 to 40,000 years ago as it has Regions of Difference (RD), H37Rv related deletions- RvD and TB specific deletions-TbD1 that are absent in the other six members (114) (fig1-6). Thus genetic polymorphism can occur in MTBC either due to a single nucleotide mutation in specific gene coding regions, or by the presence or absence of a gene or region, thereby altering the gene content of the microorganisms (fig. 1.13). Linkage Disequilibrium (LD) studies now support the theory that *M. tuberculosis* has evolved clonally (152,153). LD refers to the non-random association of alleles at two different loci, which are not independent of one another (154).

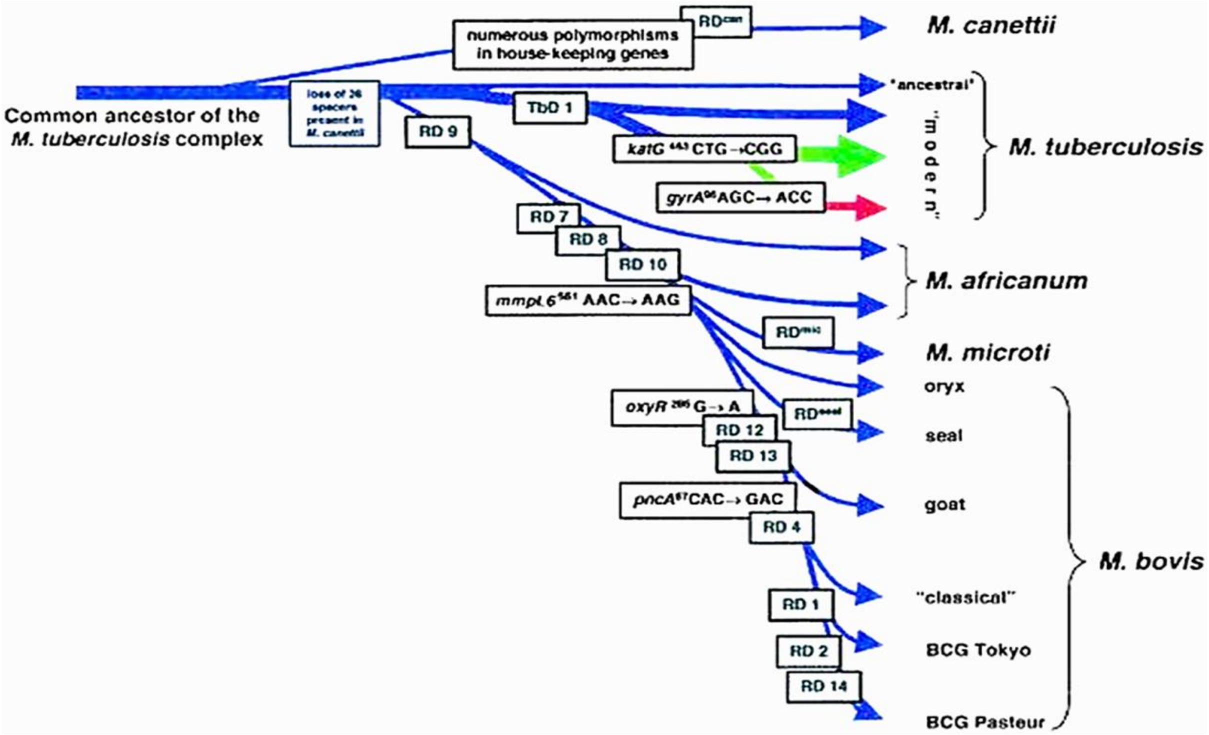


Figure 1-13 Novel evolutionary scenario of MTBC proposed by Brosch et al in 2002 based on the regions of difference (RD) and sequence polymorphisms within five genes.

Blue arrows indicate strains that are characterised as group one organisms, the green arrow is group two and the red arrow group three, as defined by $katG^{463}$ and $gyrA^{95}$ codon sequence polymorphisms as described by Sreevatsan et al., in 1997 (153) as shown in figure 1.4 (Taken from .(114))

Several clonal species of mycobacteria are not identifiable when they are isolated on solid media as they are thought to have descended from a common ancestor and MTBC fall under this category as previously discussed.

1.9.1 Major Genetic Groups of MTBC and its global distribution

Major Genetic Groups (MGG) classification is based on single nucleotide mutations or polymorphism (SNP). Generally, there are two types of SNPs synonymous and non-synonymous. The former are important to study evolution as they are not prone to divergence whereas the latter refers to translational changes that are not evolutionarily informative (155). Sreevatsan et al., in 1997 identified two SNPs; one in the catalase peroxidase (*kat*G codon 463) and one gyrase A (*gyr*A codon 95) encoding genes and classified MTBC into three Major Genetic Groups (MGGs) or Principal Genetic Groups (PGGs). Here they showed that there are three MGGs- MGG1, MGG2 and MGG3 where MGG3 is derived from MGG2 and MGG2 from MGG1. An *M. tuberculosis* precursor or common ancestor would have given rise to MGG1 (153) (fig 1-14).

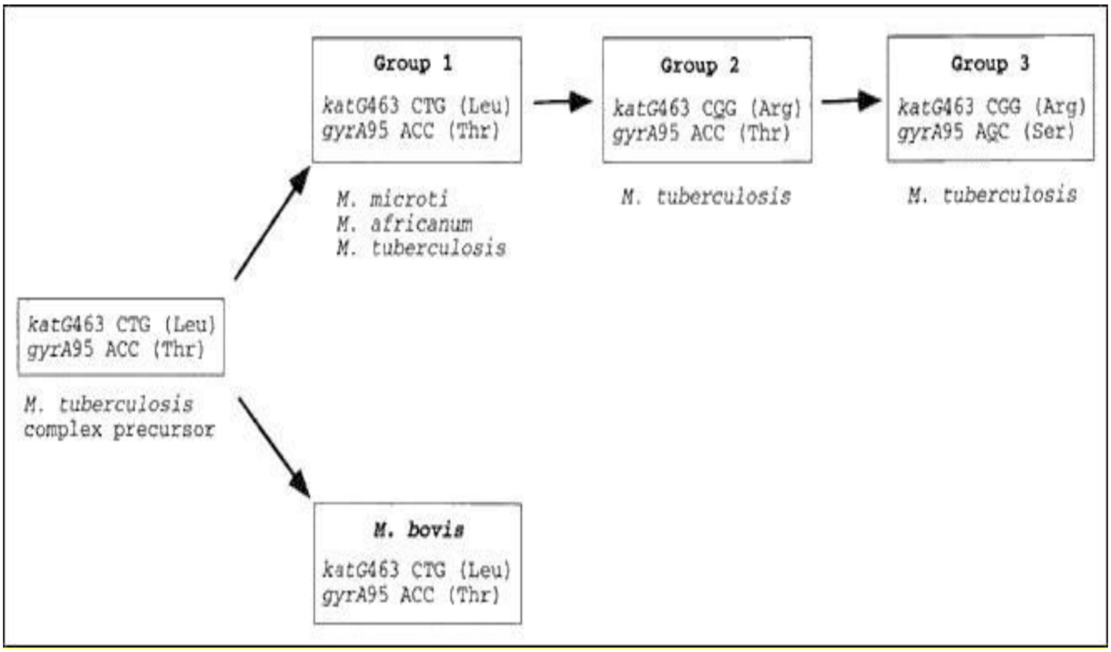


Figure 1-14 Evolutionary scenario proposed by Sreevatsan et al, in 1997 based on SNP polymorphism

(Taken from.(153))

Several other studies by Gutacker et al., 2006, Baker et al., 2004, Rad et al., 2003 have all classified the MTBC into different lineages either based on SNPs or the presence or absence of TB specific deletion regions (TbD1) (155–157). Brosch et al. in 2002 showed that the MTBC diverged from the most recent common ancestor over thousands of years by losing regions of deletion (RD) or due to large sequence polymorphisms (LSPs) in house-keeping genes. It is important to note that the TbD1 deletion is common to all ‘modern’ TB strains which include the Haarlem, Beijing and CAS spoligotype strains and that this deletion is irreversible because of the lack of DNA recombination between strains of the MTBC. The TbD1 deletion is therefore absent from ancient strains such as the East African-Indian (EAI) strains. The division of modern and ancient families (where DNA markers characterised, not the genomes themselves, are thought to be similar to those found in modern or ancient strains) is based on spoligotyping, identifying the deletions or polymorphism in the DR locus (DVR24) in *M.tuberculosis* genomes (112). Different spoligotypes or families derived by adopting this technique, based on detection of the presence or absence of 43 unique sequences in the Direct Repeat region, are given below in the table 1-6.

Table 1-6 Table showing the distribution of *M. tuberculosis* complex (MTBC) strains with respect to the Principal Genetic Groups (PGGs) and the TB spoligotypes families / lineages

Principal Genetic Groups	Mycobacteria	TB Spoligotype family / lineage
PGG1	*M. microti, M. africanum, M. tuberculosis* and *M. bovis*	Central Asian Strain (CAS) Beijing
PGG2	*M. tuberculosis*	S X T-Uganda Haarlem Latin American Mediterranean (LAM)
PGG3	*M. tuberculosis*	T

Due to the inclusion of various sub-species of MTBC (4) including the classification of TB genetic lineages according to the geographical migration (158), the evolutionary scenario originally proposed by Sreevatsan et al. in 1997 and Brosch et al. in 2002 has been combined as shown in fig 1.15 below.

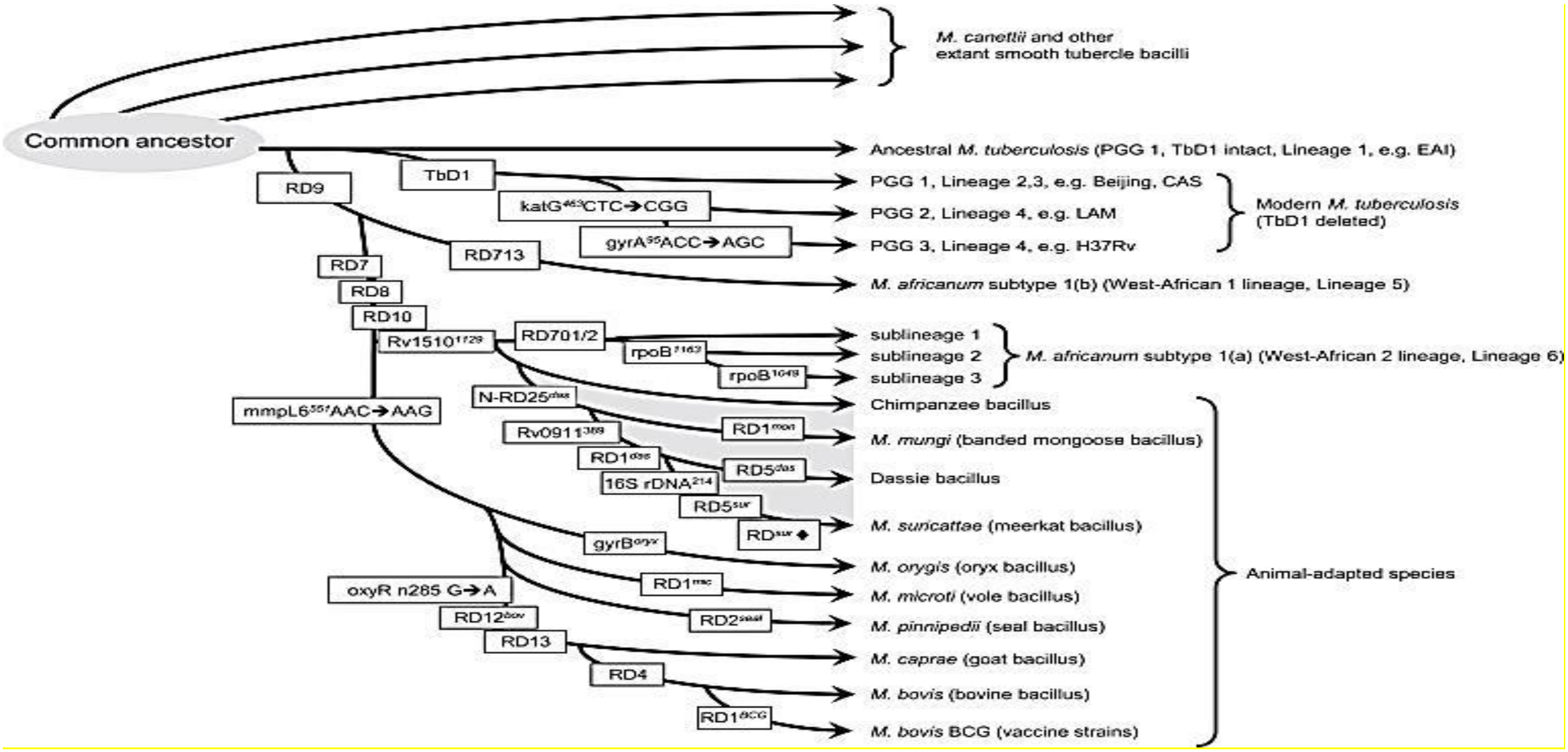


Figure 1-15 Schematic representation of the proposed phylogenetic relationships between members of MTBC.

The phylogeny is based on presence or absence of Region of Difference (RD) (114), SNPs (153) and Large Sequence Polymorphism (LSP) (159). (Taken from (4)).

1.9.2 Global distribution of TB genetic lineages including Co-Evolution of TB and Humans

It has been suggested that *M.tuberculosis* might have co-evolved with the human host and have followed the ‘out-of-and-back-to-Africa’ evolutionary scenario (159–161). In 2006, Gagneux and his colleagues addressed the co-evolutionary scenario of both human and TB populations for the first time and also proposed six phylogeographical lineages (fig 1-16) (158).

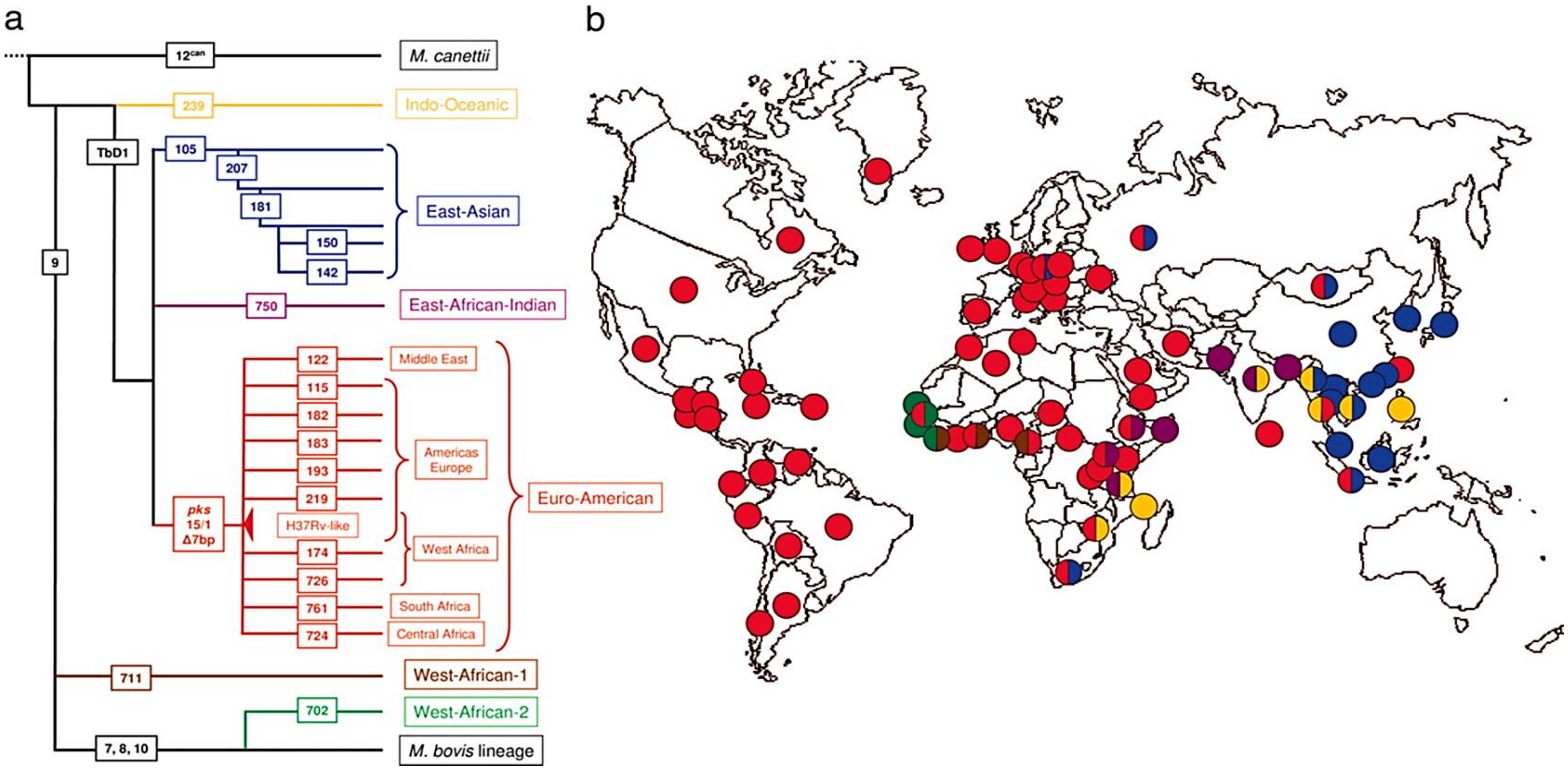


Figure 1-16 The phylogeographical distribution of *M. tuberculosis* and its lineages

(*a*) LSPs define a global phylogeny for *M. tuberculosis*. The names of the lineage-defining LSPs or regions of difference are shown in rectangles. The geographic regions associated with specific lineages are indicated. (*b*) The six main lineages of *M. tuberculosis* are geographically structured. Each dot corresponds to 1 of 80 countries represented in the global strain collection. The colours of the dots relate to the six main lineages defined in (a) and indicate the dominant lineage(s) in the respective countries (Taken from (158))

1.9.3 Molecular Clocks in the evolution of *M.tuberculosis*

Using different molecular markers (SNPs, RDs, TbD1 and spoligotypes), a time line of evolutionary events can be generated that strongly suggests that the currently circulating MTBC isolates might have originated from a common ancestor as shown in figure 1-17 below (162) as discussed in section 1.11 and shown in fig 1-13 by Brosch et al (114).

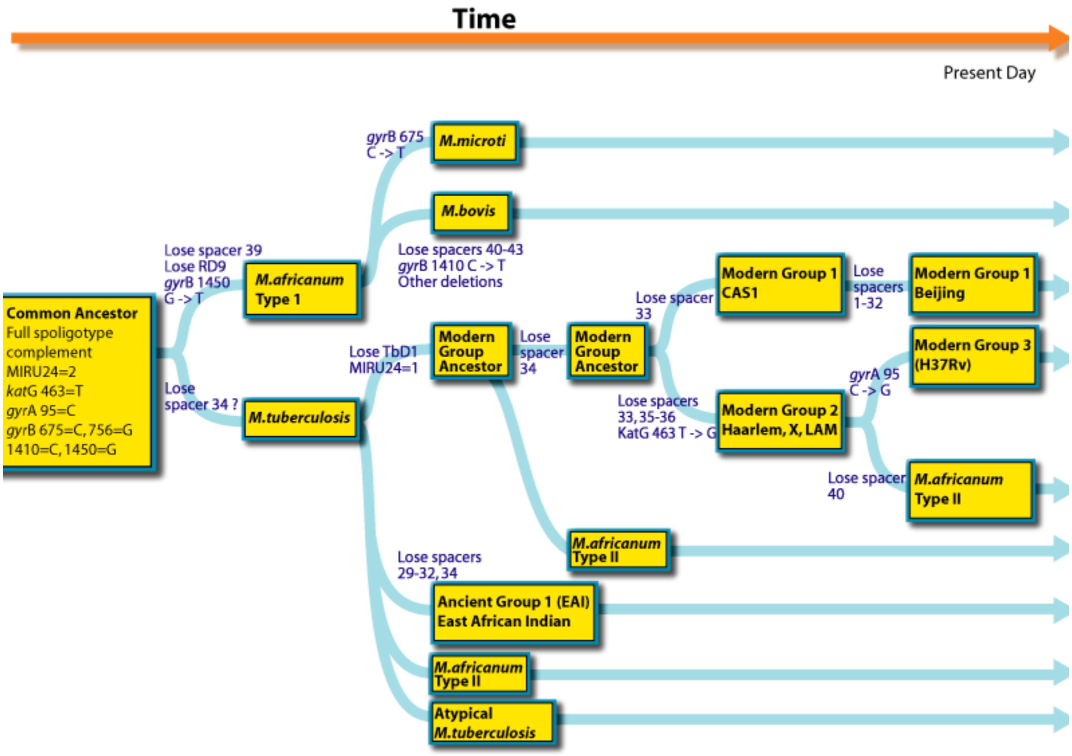


Figure 1-17 Timeline of evolutionary events in *Mycobacterium tuberculosis*

(Taken from (162))

Over recent years, data have accumulated suggesting that the different MTBC lineages might have adapted to different human populations and thus showing predominance in certain regions like CAS in Northern India (163) and Beijing in Tibet and China (164). Not only does MTBC exhibit a global biogeographic population structure, but the associations between the particular MTBC lineages and human populations are maintained in cosmopolitan settings where human populations and their associated MTBC strains experience at least some degree of intermingling (161,165,166). Sequencing six housekeeping genes (*katG, gyrB, gyrA, rpoB, hsp65* and *sodA*) and the complete 16SrRNA gene in smooth tubercle bacilli suggests that *M. tuberculosis* might have descended 3 million years ago (167). Also, ancient DNA studies (160) suggest that MTBC predates the arrival of the human population (fig 1-18), thus showing the adaptability of this ancient microorganism.

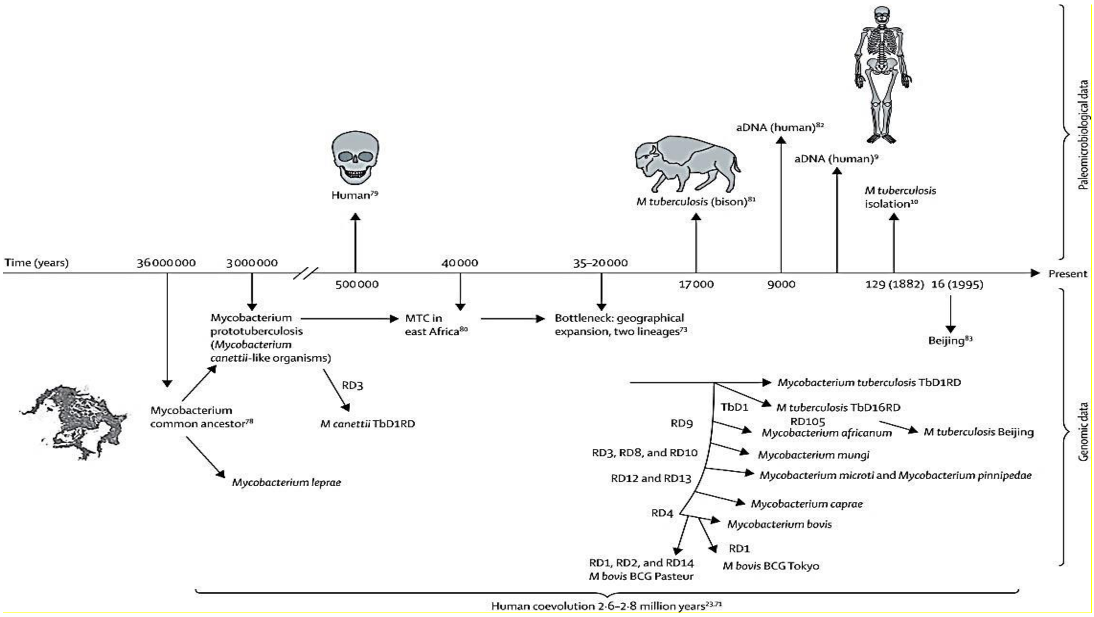


Figure 1-18 Evolution of *Mycobacterium tuberculosis* complex along with human population

(Taken from(160)).

1.9.4 Importance of genetic polymorphism

Different markers have been used to study the phylogenetic relationships between the common ancestor and the currently circulating TB genetic lineages. As IS*6110* is one of the earliest seen polymorphic changes in terms of evolution of TB, they could theoretically play an important role in the evolution of different TB genetic lineages. IS*6110* insertions occur mainly in multi-gene families such as the PPE gene family and they could be beneficial for the organism to evade immunity (74) and cause disease (168). In addition, several studies have suggested that drug resistance, formation of LSPs and high frequency of SNPs were noticed in regions flanked by IS*6110* (169). Thus, IS*6110* transposition could be essential in the evolution of the *M.tuberculosis* genome.

1.10 Molecular Epidemiology of TB in resource poor settings

Among the estimated 9.4 million new TB cases reported in 2008, 55% and 30% of the cases were reported in resource poor countries in Asia and Africa respectively (170). Depending on the availability of instruments and reagents, the techniques used to study the molecular epidemiology of TB are different in different countries. Epidemiological studies are however limited when compared to developed countries due to financial constraints in these areas despite the fact that tuberculosis is endemic.

1.11 Currently used epidemiological tools, its suitability and constraints

The tools used in a resource-limited country mainly depend on the country's ability to afford equipment required for different technologies, their maintenance and the availability of the skilled force to perform the techniques. Resource poor countries may need improved access to diagnosis and care (171). Without an effective surveillance strategy, TB control measures could be difficult to be monitored. So, routine surveillance methods will contribute in assessing drug resistance strains and monitoring. Within their limits these settings continue to contribute to the understanding of the molecular epidemiology of TB using the tools available to them including the most commonly used techniques IS*6110* RFLP, MIRU-VNTR and spoligotyping.

1.12 Nepal TB epidemiology

Though the global TB burden has been falling globally in the last few years due to the introduction of the STOP TB programme by the WHO (1), the incidence of HIV-associated TB is high in Myanmar, Thailand, India and Indonesia. Nepal used to be one of the five countries listed in SE Asia region with a high burden of TB But the new global TB report from WHO (172) has excluded Nepal from the list of 30 high burden countries due to the effective STOP TB policy since 2006. But tuberculosis still remains a public health problem and ranked as the sixth leading cause of death among the top 20 causes of death in Nepal according to the National Tuberculosis programme survey carried out in 2014 (173).

1.12.1 TB in Nepal

Nepal has 147,181 square kilometres of land mass area with a population of 29 million people and is interlocked between the Indian subcontinent on three sides and China’s Tibet to the North. Nepal lies on the major trade route between India and China (see figure.1-19) and has an 1200-km long controlled border with China in the north and an 1700-km long open border with India in the east, west and the south (174). It has been noted that migration across these international borders take place for reasons like work, study, tourism, religious purposes like pilgrimages and cultural exchanges like marriages. Kathmandu, the capital city of Nepal is highly populated with a high number of TB cases according to the Nepal's National Tuberculosis programme.

Figure 1-19 A map of Nepal showing the topography of Nepal interlocked between India (in the south, east and west of Nepal) and China (in the north)

(Taken from http://www.echoway.org/graphic/oupartir/cartespays/nepal.gif)

In 2014, the WHO estimated that 4600 (2100-7500) people died from TB in Nepal. Even though MDR-TB levels are low (2.2%) in new cases (high in retreatment cases-15.4%), Drug-resistant TB (DR-TB) is still a major public health concern. The epidemiological indicators of the TB burden in Nepal according to the report published in 2014 by National Tuberculosis Programme is summarised below in the table 1-7.

Table 1-7 Estimated TB burden in Nepal in 2014

(Adapted from NTP, Nepal 2014 report)

Epidemiology Indicators	Estimated patient cases (Range)
Prevalence	59 (27-100) x 1000
Prevalence rate	211 (99-365)
Incidence	43 (39-49) x 1000
Incidence rate	156 (139-178)
Cases notified to NTP	37025
Annual case notification rate	136 / 100,000
HIV negative / TB positive cases	4.6 (2.1-7.5) x1000
HIV positive / TB positive cases	0.32 (0.17- 0.51) x1000

1.12.2 Current diagnostic procedures in Nepal

Sputum smear microscopy is used as the main method to detect TB and monitor treatment responses in patients. Culturing the bacteria and phenotypic Drug Susceptibility Tests (DSTs) are the two principal procedures carried out to monitor new and relapsed cases within the 581 microscopy centres, two solid culture laboratory including capacity for first line drug (FLD) susceptibility testing (DST), first line probe assay (LPA), one liquid culture and one second line drug (SLD) DST facility and 26 GeneXpert centres spread all over this mountainous country. This is in addition to two other culture and DST facilities (NTC and GENETUP) functioning in Kathmandu valley.

1.12.3 Constraints in diagnosis of TB in Nepal

According to the NTCP's annual report, the challenges faced by the authorities to control TB in Nepal are mainly due to the increase in MDR and XDR cases, low case notification by different districts, insufficient supervision, monitoring and evaluation of TB cases and lack of skilled labour (173). In a recent survey in 2014, it has been found that nearly 9.3% of new patients are resistant to at least one drug. Also, resistance to fluoroquinolones (26.4%) has led to a heavy burden of pre-XDR and XDR-TB among MDR-TB cases (8% of the cases were found to be XDR among MDR pateints). It has been speculated that it might be due to the fact that fluroquinolones could be purchased by the patients over the counter easily and also its unregulated use. Case Notification Rate (CNR) in 2015 was 123/ 100,000 a drop of 11% compared to the previous survey in 2010 (173). Among different regions, there has been a drop in the CNR in the Eastern and the central regions of Nepal. So, an efficient, cost-effective and rapid epidemiological tool that can be useful in these settings was needed and thus this research project aimed to develop a genomic tool box that can be easily used in these settings along with other techniques.

1.13 Aims and Objectives

The aim of my PhD thesis is to develop a genomic tool box to aid epidemiological investigations using a novel IS*6110* FAFLP sampling method for application in resource poor settings.

1.14 Hypothesis

Characterisation of IS*6110* insertion sites in TB genomes of strains from poorly studied regions will give insight into the geographical distribution of disease and their relationships with different lineages thereby showing that IS*6110* is not only a typing tool but also an effective evolutionary marker.

The objectives of this thesis are as follows:

- **Chapter 1**: To review the TB literature in detail focusing on the suitability of the technique and the sample population.
- **Chapter 2**: To describe IS*6110* FAFLP methodology in addition to the reagents needed for the technique.
- **Chapter 3**: To optimise and develop the IS*6110* FAFLP methodology using H37Rv and Nepal strains.
- **Chapter 4**: To map the IS*6110* insertion sites in the *M. tuberculosis* H37Rv genome and to rapidly detect TB genetic lineages using IS*6110* FAFLP derived PCR.
- **Chapter 5**: To classify the TB samples from Nepal into their different TB genetic lineages.
- **Chapter 6**: To understand the Rifampicin resistance status in the TB samples from Nepal.

- **Chapter 7**: To contextualise that IS*6110* FAFLP as a genomic mapping tool and discuss its suitability as a method to understand the phylogenetic relationships between TB genetic lineages and further indicate that IS*6110* is a valuable tool to study the epidemiology and evolution of TB.

Chapter 2 General Materials and Methods

2.1 Bacterial samples

2.1.1 Sample collection centres

Sputum samples from consecutive new TB patients were collected over one year between 2007 and 2008 and cultured alongside routine diagnostic testing from two Nepalese tuberculosis reference centres located in the Kathmandu valley: the National Tuberculosis Centre (NTC) and the German Nepal Tuberculosis Project (GENETUP) (fig 2.1).

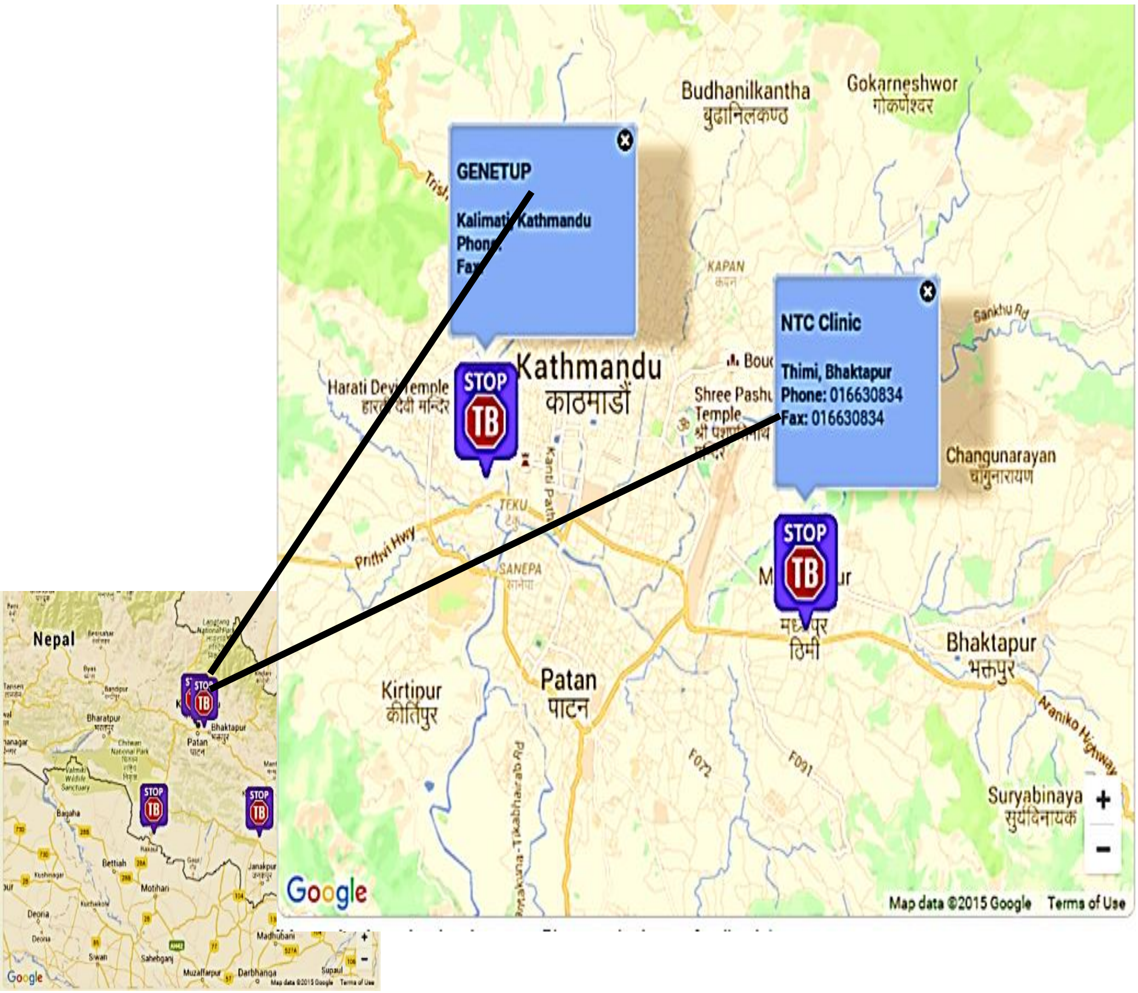


Figure 2-1 Geographical location of the two Nepal TB treatment centres, GENETUP and NTC, where samples were collected

(Adapted from http://www.nepalntp.gov.np/index.php?view=location)

Bacterial genomic DNA from the strains isolated from the above samples were extracted at the National Mycobacterial Research Laboratory (NMRL) in Nepal using Cetyltrimethyl Ammonium Bromide (CTAB)(175).

2.1.2 Ethics

Nepal Health Research Council (NHRC), the Ethics Approval Committee approved the research projects ongoing at MRL in general between 2006 and 2008 when these samples were collected. During this time, the specific procedures including consent were approved by the Head of MRL, MRL Lab Manager, the Anandaban Medical Director, faculty of Tribhuvan University and Directors of both Tuberculosis centres where samples were collected. As sputum samples were classified as non-invasive samples during collection, written consents were not emphasised and only verbal consents were taken but not documented from the patients who attended the hospitals primarily for TB treatment. The anonymized DNA samples were then transported to Public Health England (PHE), Colindale in 2008 and this particular study started in 2011 and was solely conducted in Public Health England, Colindale.

2.1.3 IS*6110* Fluorescent Amplified Fragment Length Polymorphism (FAFLP)

This method published by Thorne et al. (146) was optimised and developed as discussed in chapter 3. The method comprised of four major steps: Digestion of genomic DNA, Ligation of adapters, PCR and Capillary Electrophoresis (Figure 2-2). All reagents used in this study were risk assessed locally as the Genomic Services Development Unit (GSDU) 2D01 lab where this research was carried out already uses the FAFLP procedure for the day-to-day service work.

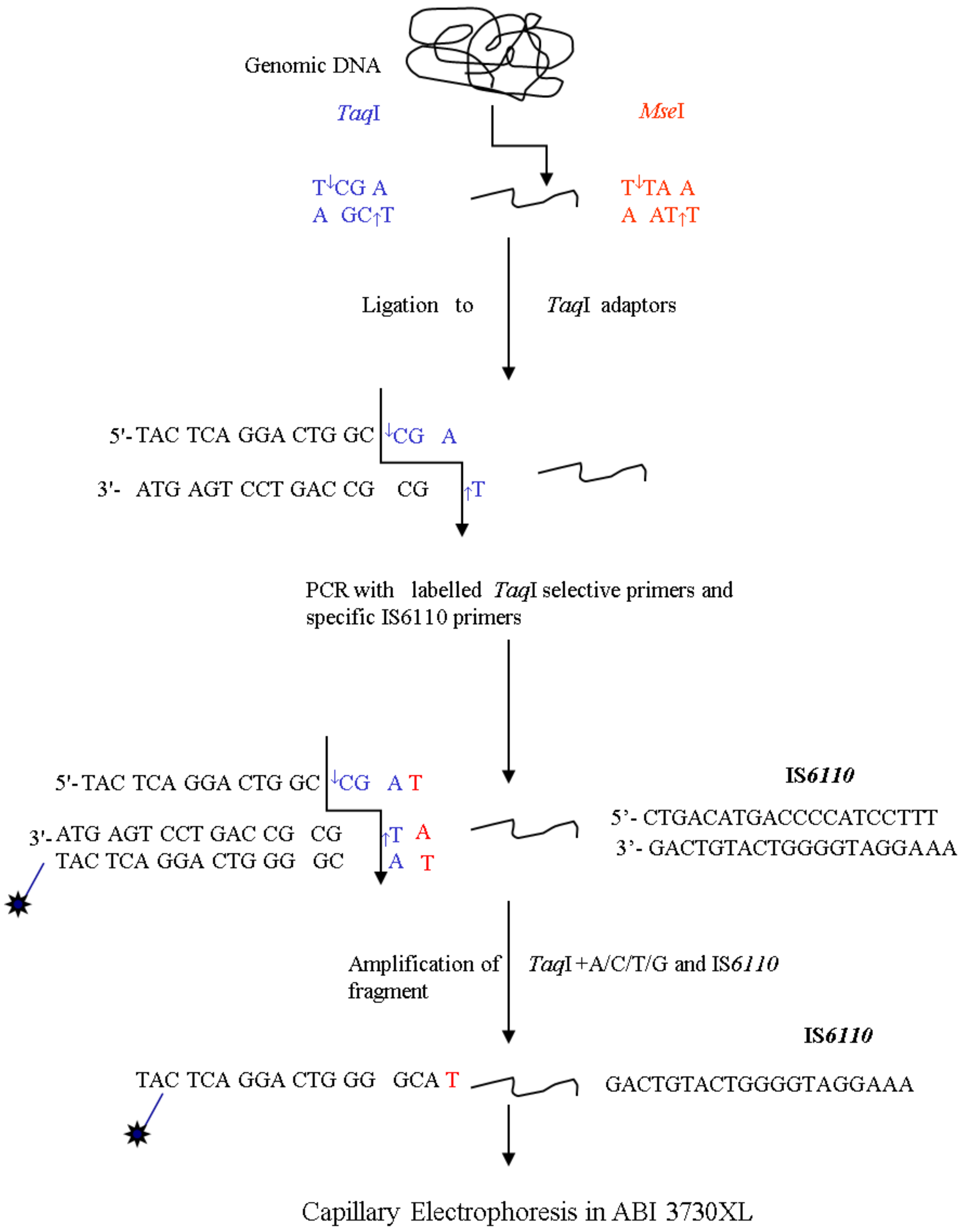


Figure 2-2 Schematic representation of the IS*6110* FAFLP methodology.

2.1.4 Genomic DNA Restriction Endonuclease digestion

Reagents:

1. *Mse*I (New England Biolabs (NEB), UK)
2. *Taq*I (NEB, UK)
3. 1x *Mse*I buffer II (NEB)
4. Bovine Serum Albumin (BSA) , (NEB)
5. Ribonuclease A (RNase A) (Sigma Aldrich, UK)
6. 1M proline (Sigma Aldrich, UK)
7. 2mM betaine (Sigma Aldrich, UK)
8. Molecular Biology Grade water (Fisher Scientific, UK).

2.1.4.1 Protocol:

Five µl of 200- 500 ng of bacterial genomic DNA quantified using Qubit (Thermofisher, UK) was digested with 5 U of *Mse*I (New England Biolabs, UK) in a total reaction volume of 20µl containing 1x *Mse*I buffer, 0.1mg/ml of bovine serum albumin, 0.5mg/ml of RNase A, 1M of L-proline and 2mM of betaine at 37°C for 2 h to release the DNA ready for a final digestion with 10U of *Taq*I (New England Biolabs, UK) at 65°C for 3 h and the reaction was inactivated for endonucleases at 80°C for 10 min.

2.1.5 Ligation of digested DNA to double stranded adapters

2.1.5.1 Reagents:

1. T_4 DNA Ligase (NEB, UK)

2. 2 x T_4 DNA Ligase buffer (NEB, UK)

3. Double stranded *Taq* I adaptor (Eurofins, UK)

2.1.5.2 Protocol:

Double stranded oligonucleotide adaptors specific to *Taq*I site (5’-TACTCAGGACTGGC) were ligated to the double- digested DNA in a total reaction volume of 25µl containing 40U of T_4 DNA ligase (New England Biolabs, UK), 0.2µM of double stranded *Taq*I adaptor and 2x T4 ligase buffer (NEB, UK). The reaction mixture was then incubated at 12°C overnight followed by heating at 65°C for 10 minutes to inactivate the ligase and later stored at -20°C.

2.1.6 PCR amplification of fragments

2.1.6.1 Reagents:

1. 10 x PCR reaction buffer (Thermofisher, UK)

2. 50mM MgCl2 (Thermofisher, UK)

3. 10mM dNTPs (Thermofisher, UK)

4. 5µM *Taq*I primer (Eurofins, UK)

2.1.6.2 Preparation of 10µM of 4-dye labelled *TaqI* primer

10µM of each selective *Taq*I primer labelled with 6-FAM (blue), PET (red), VIC (green) and NED (Black) fluorescent dyes at the 5’ end and a selective base either A or C or G or T nucleotide respectively according to the dye listed above were made from 100µM stocks with molecular biology grade water. 10µl of each 10uM primer i.e. 40µl in total were then mixed with 60ul of water to make up 10uM *Taq*I primer stock.

5. 5uM IS*6110* specific primer (Eurofins, UK)
6. Molecular Biology Grade water (Fisher Scientific, UK)
7. Recombinant *Taq* DNA Polymerase (Thermofisher, UK)

2.1.6.3 Protocol:

Four labelled adapter-specific forward primers, each with a single unique selective nucleotide (A / C / T/G) at the 3’ end which extends into the unknown genomic DNA sequence of the fragment, together with an IS*6110* specific reverse primer, were used to amplify the fragments. In a total volume of 20 µl, 1µl of the adaptor ligated DNA was added to the reaction containing 1X reaction buffer (Invitrogen, UK), 1.5 mM $MgCl_2$ (Invitrogen, UK), 0.2 mM dNTPs (Invitrogen, UK), 1 µM of labelled selective *Taq***I** forward primer (5’-CGATGAGTCCTGACCG**A**/**C**/**T**/**G- (**see table 2-1), 1 µM of IS*6110* reverse primer (5’- CTGACATGACCCCATCCTTT) and 1U of recombinant *Taq* polymerase (Invitrogen, UK). The following PCR conditions were carried out in a Veriti thermocycler (Applied Biosystems, UK): 94°C for 15 min followed by 35 cycles of 94°C for 20 s, 66°C for 30 s and 72°C for 2 min with the 66°C annealing temperature reducing by 1°C every cycle for nine cycles and the last 25 cycles at 56°C. Finally, an extension of 72°C for 60 min was carried out before further manipulations.

Table 2-1 List of *Taq*I fluorescently labelled selective primers

Fluorescent labelled selective primer	**Primer Sequence (selective base** (135,147)
1) *Taq*I- **A** (labelled with 6-FAM dye)	5'- CGATGAGTCCTGACCGA(**A)**-3'
2) *Taq*I- **C** (labelled with PET dye)	5'- CGATGAGTCCTGACCGA(**C)**-3'
3) *Taq*I- **T** (labelled with NED dye)	5'- CGATGAGTCCTGACCGA(**T)**-3'
4) *Taq*I- **G** (labelled with VIC dye)	5'- CGATGAGTCCTGACCGA(**G)**-3'

2.1.7 Capillary Electrophoresis of fragments using ABI3730xl Genetic Analyser

2.1.7.1 Reagents:

1. Hi_Di formamide (Applied Biosystems, UK)

2. Genescan Liz600 sizing standard (Applied Biosystems, UK)

2.1.7.2 Protocol:

One microlitre of the IS*6110* FAFLP PCR reaction from the PCR step was mixed with 10µl of Hi-Di formamide (Applied Biosystems, UK) and 0.5 µl of the Genescan Liz600 sizing standard (Applied Biosystems, UK). The reaction was denatured at 95° C for 5 min before running the reaction on the ABI genetic analyser 3730xl. The run module settings are given in the table 2.2 below:

Table 2-2 Run Module settings for capillary electrophoresis of FAFLP fragments

Run parameters	**Actual setting**	**Range**
Oven temperature	63°C	18-70°C
Buffer temperature	35°C	30-35°C
Prerun voltage	15.0 kV	0-15 kV
Prerun Time	180 sec	1-1800 sec
Injection voltage	1.6 kV	0-15 kV
Injection time	15 sec	1-90 sec
First readout time	200 ms	100-16000 ms
Second readout time	200 ms	100-16000 ms
Run Voltage	15 kV	0-15 kV
Voltage no. of steps	10	0-100 steps

Voltage step interval	20 sec	0-180 sec
Voltage tolerance	0.6 kV	0-6.0 kV
Current stability	30 µA	0-2000 µA
Ramp Delay	1 sec	1-1800 sec
Data Delay	500 sec	1-1800 sec
Run Time	2700 sec	300-14000 sec

Chapter 3 Optimisation and Development of Insertion Element IS*6110* Fluorescent Amplified Fragment Length Polymorphism (FAFLP) PCR

.

3.1 Introduction

Fluorescent Amplified Fragment Length Polymorphism (FAFLP) has been widely used as a strain typing tool for several bacteria like *Listeria monocytogenes* (134), *Legionella pneumophila* (145,176) and *Salmonella enteritidis* (142). Using IS*6110* as a biomarker, the global phylogeny of MTBC has been studied using FAFLP (135).

In the era of whole genome sequencing, where repeat elements are difficult to sequence with short read technologies (177), and in the absence of an alternative approach to their characterisation in poorer resource settings, a simple but effective technique would enable the facilitation of outbreak samples by not only strain typing the bacterial strains but also to help in the classification of TB genetic lineages in any geographical location. A rapid and simple method of insertion site mapping using IS*6110* fluorescent amplified fragment length polymorphism (FAFLP) PCR was developed using fluorescent labelled primers (147). Borrell et al. first used four different fluorescent labelled primers with the selective base at the 3' end of the *Taq*I primer for increased resolution to differentiate between epidemiologically unrelated TB strains. Thorne et al. used fluorescent labelled primers to further distinguish the fragments (135,146). To suit the demands of a resource poor setting the technique was modified further by utilising the resolution of the fragments derived by this technique, and to develop a lineage specific PCR as described in chapter 4.

3.1.1 Aims

The main aim of this chapter was to modify, optimise and standardise the IS*6110* FAFLP PCR methodology so that it could be used to map and classify *M. tuberculosis* samples into different lineages.

3.1.2 Objectives

- To test the IS*6110* FAFLP methodology using DNA from different strains of *M. tuberculosis* and different PCR reagents/conditions.

3.1.3 IS*6110* FAFLP Methodology

The technique was optimised so that the results obtained could be easily identified by the presence or absence of the fragments/peaks and by their respective colours as shown in the figure 3-1. Each fragment indicates the point of insertion of IS*6110* in the bacterial genome and also the number of fragments indicates the number of IS*6110* insertion sites i.e. copy number of IS*6110* in the genome. The PCR amplified fragments were run on ABI 3730XL (Thermofisher, U.K.) and visualised using software called Peak Scanner (Thermofisher, U.K.).

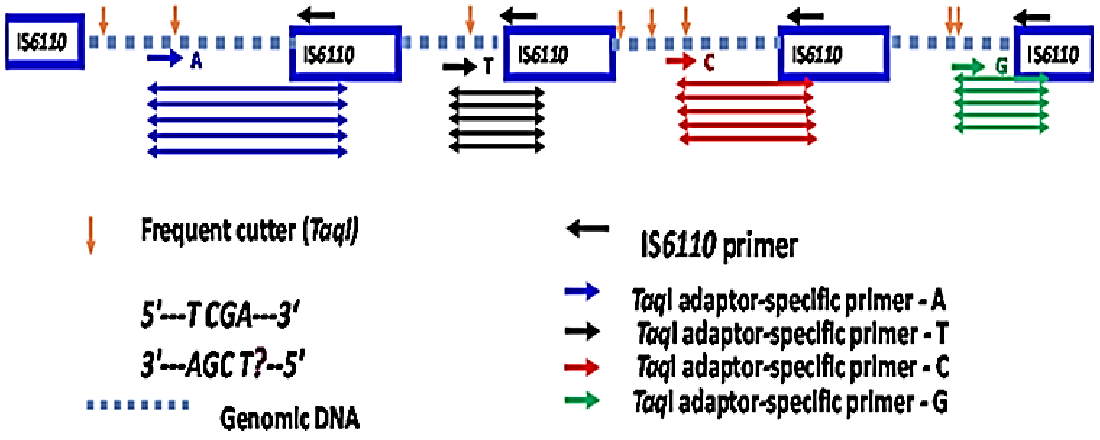


Figure 3-1 Pictorial representation of the IS*6110* FAFLP PCR methodology

(Coloured fragments are amplified using uniquely labelled adapter specific primers. For example, blue coloured fragments are amplified if the primers are labelled with 6-FAM fluorescein 'blue' dye, red coloured fragments with PET dye, green coloured fragments with VIC dye and black coloured fragments with NED dye).

3.2 Materials and Methods

The process to optimise the IS*6110* FAFLP PCR methodology consisted of testing the DNA samples of adapter ligated H37Rv DNA from previous study by Kremer et al.(178), H37Rv and Nepal TB strains with different polymerases and optimising the PCR conditions. Approximately 200-500ng of bacterial genomic DNA was digested with 5U of *Mse*I (New England Biolabs, UK) in a total reaction volume of 20µl containing 1x *Mse*I buffer, 0.1mg/ml of BSA, 0.5mg/ml of RNase A, 1M of L-Proline and 2mM of Betaine at 37°C for 2h to release the DNA ready for a final digestion with 10U of *Taq*I (New England Biolabs, UK) at 65°C for 3 h and the reaction was inactivated for endonucleases at 80°C for 10 min. Double stranded oligonucleotide adaptors specific to *Taq*I site (5'- TACTCAGGACTGGC) were then ligated to the double- digested DNA in a total reaction volume of 25µl containing 40U of T_4 DNA ligase (New England Biolabs, UK), 0.2µM of double stranded *Taq*I adaptor and 2x T4 ligase buffer (NEB, UK). The reaction mixture was then incubated at 12°C for 17 h followed by heating at 65°C for 10 minutes to inactivate the ligase and later stored at -20°C. In a total volume of 25µl, 2.5µl of the DNA sample was mixed with 10x *Pfx* reaction buffer, 50mM $MgSO_4$, 10mM dNTP, 1µM of *Taq*I primer, 1µM of IS*6110* specific primer and 1U *PfX* polymerase/ 1µl of Hotstar *Taq* plus mastermix or 10µl of My *Taq* Premix or 1U of recombinant *Taq* polymerase. The following PCR conditions were carried out in a Veriti thermocycler (Thermofisher, U.K.): 94°C for 15 min followed by 35 cycles of 94°C for 20 s, 66°C for 30 s and 72°C for 2 min with the 66°C annealing temperature reducing by 1°C every cycle for nine cycles and the last 25 cycles at 56°C. Finally, an extension of 60°C for 30 min was carried out before further manipulation. When there were no peaks or there was an anomaly in the electropherogram such as amplification artefacts and double peaks, then the PCR extension was increased from 60°C for 30 min to 72°C for 30min and a final adjustment was carried out from 72°C for 30min to 72°C for 60min. The criteria for the peak characteristics seen in an electropherogram using Peak Scanner software were as follows:

a. The fragment should be a single peak (Double peaks should be avoided). Two peaks over 0.5 base pair difference between them will be considered as two separate fragments.

b. The fragment could be identified by the colour of the fluorescently labelled dye, either blue for FAM or red for PET or green for VIC or black for NED according to the dye incorporated.

c. The peak should be a true peak shape and size (artefacts were discarded). A difference of ± 0.5bp in fragment sizes was considered as non-identical fragments or peaks.

d. The relative fluorescent units (rfu) on the X-axis had been assigned a cut-off of 500rfu.

Some of the well-characterised artefacts are as follows:

a. Stutter peaks: Stutter artefacts are observed as multiple peaks preceding the true allele peak. The number of peaks and their intensities are proportional to the length of the repeat and the number of repeats in the PCR product (Applied Biosystems Fact sheet).

b. Plus A additions or Double peaks: Plus A additions, caused by incomplete A nucleotide addition, also increase the complexity of the peak pattern, which makes recognition of true allele peaks more difficult (Applied Biosystems Fact sheet).

c. Bleed through or Pull-up peaks: A Bleed-through or pull-up is a typical problem in multichannel capillary electrophoresis, which is caused by fluorescent dye cross-talk. A high signal in one fluorescent channel thereby gives rise to a pull-up peak in another channel (Applied Maths Inc., Belgium).

The step by step procedure to achieve the standardisation of IS*6110* FAFLP PCR assay is given in the form of flowchart below in the figure 3-2.

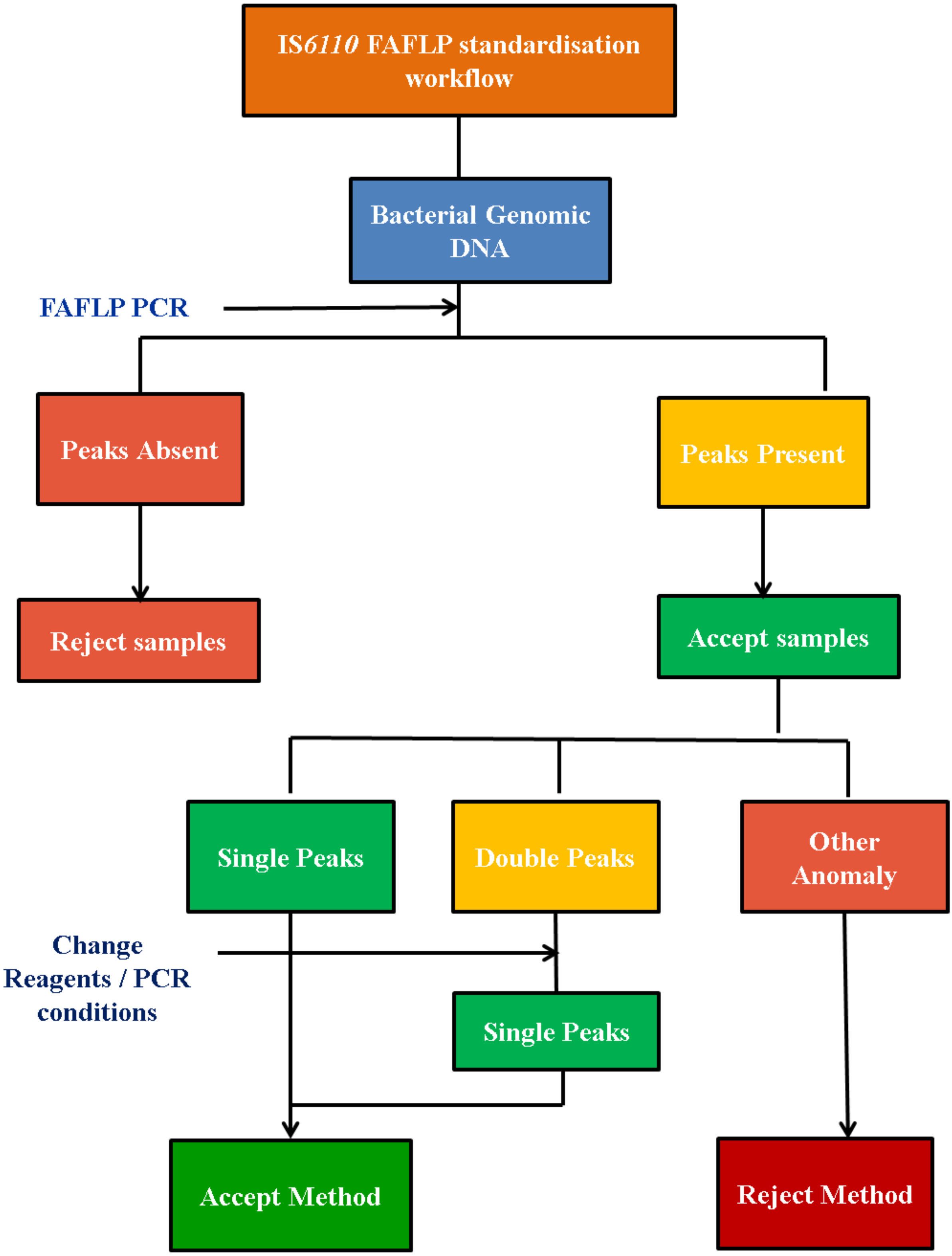


Figure 3-2 Step-by-step procedure for optimising and developing the IS*6110* FAFLP methodology

3.3 Results

The IS*6110* FAFLP PCR methodology was standardised by testing different DNA polymerases from different PCR kits and by changing the PCR extension step as shown in figure 3-3 using adapter ligated H37Rv DNA from a previous study by Kremer et al. (178), H37Rv and Nepal TB DNA samples. The sequential procedure undertaken to achieve this is described below in the form of a flowchart.

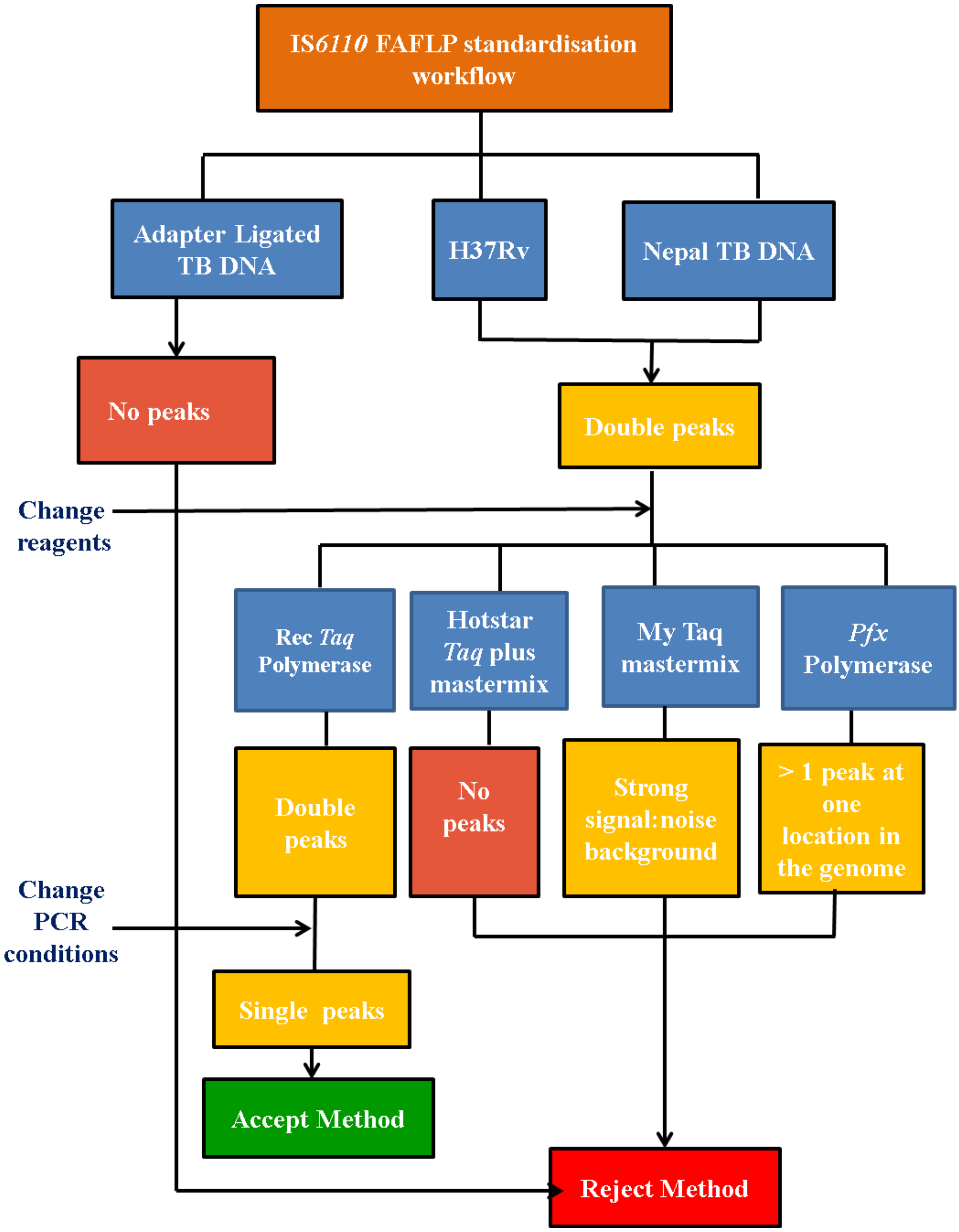


Figure 3-3 Step-by-step procedure of standardising IS*6110* FAFLP methodology including the accepted methodology

3.3.1 Amplification of DNA fragments

Amplification of the DNA fragments was first checked for the adapter ligated DNA by the presence or absence of peaks (fragments) in the PeakScanner analysis software. Absence of peaks (figure 3-4) except for the presence of LIZ600 indicated that either poor quality DNA or absence of DNA.

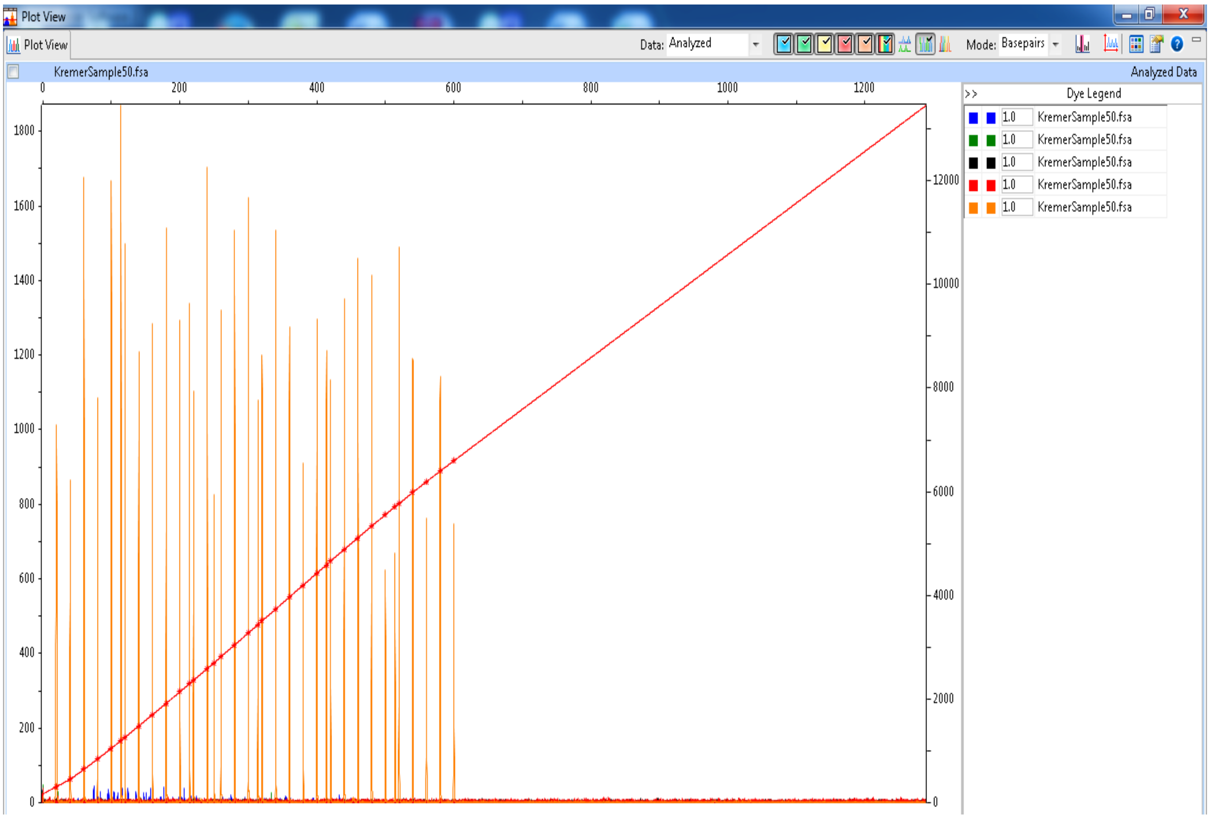


Figure 3-4 Example of a Peakscanner screenshot showing the absence of amplification of fragments except for Liz600 marker in orange colour.

3.3.2 Validation of the PCR reagents

Different Taq polymerases (*PfX* polymerase, Hotstar *Taq* plus mastermix, My *Taq* Premix and recombinant *Taq* polymerase) were subjected on both H37Rv and Nepal TB DNA and the results were as follows.

3.3.2.1 Hotstar *Taq* Plus mastermix (Qiagen, U.K.)

The fragments were not amplified except for the presence of the visualisation marker Liz600 using this mastermix and so the next reagent was tested.

3.3.2.2 My*Taq* PCR premix (Bioline, U.K.)

Using this mastermix, strong amplification occurred but the interpretation was difficult as seen in figure 3-5 due to the strong signal intensity of more than 28000 relative fluorescent units (rfu).

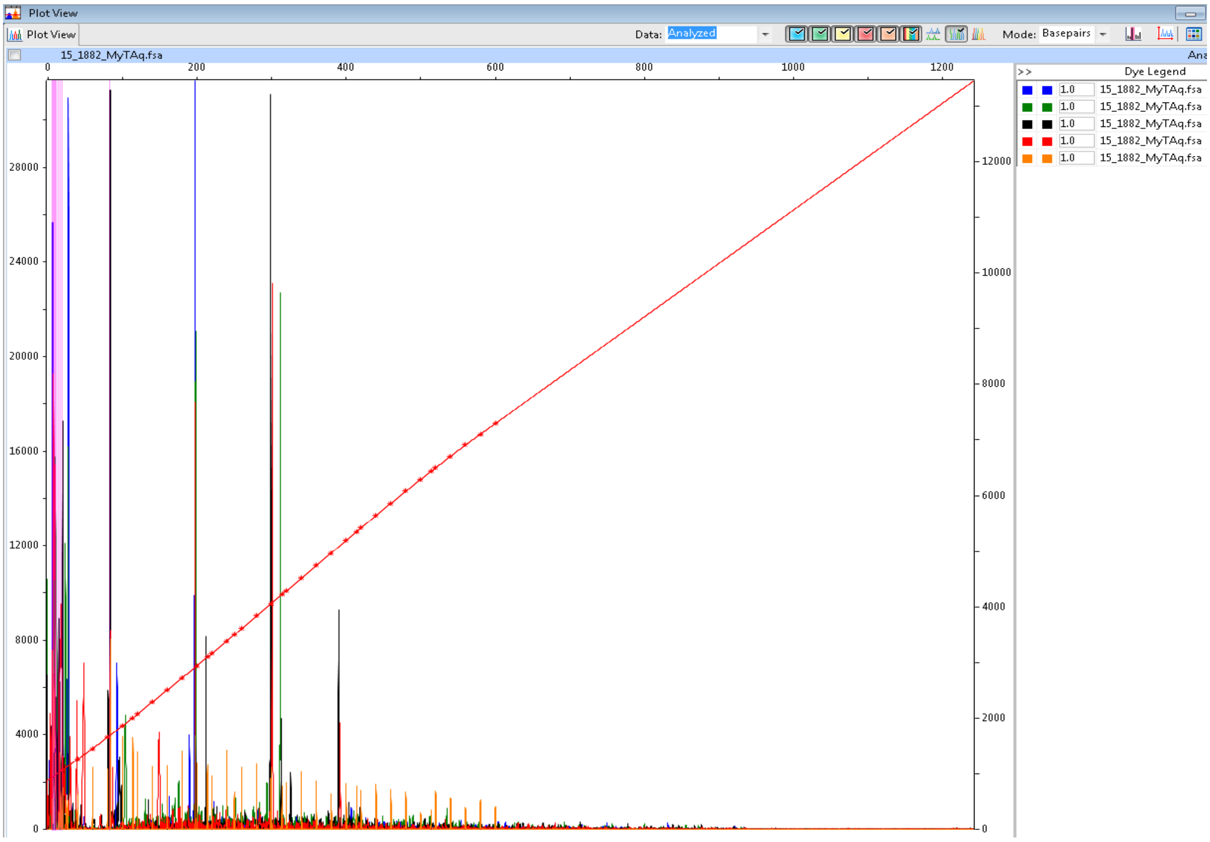


Figure 3-5 Example of a Peakscanner screenshot showing strong signals of more than 28000 rfu when My*Taq* mastermix was used

3.3.2.3 Platinum *Pfx* proof-reading polymerase (Invitrogen, U.K.)

Here, the signal intensity of the samples under analysis was optimum but the interpretation of the data became difficult as there was more than one coloured fragment at any single position as seen in figures 3-6 and 3-7.

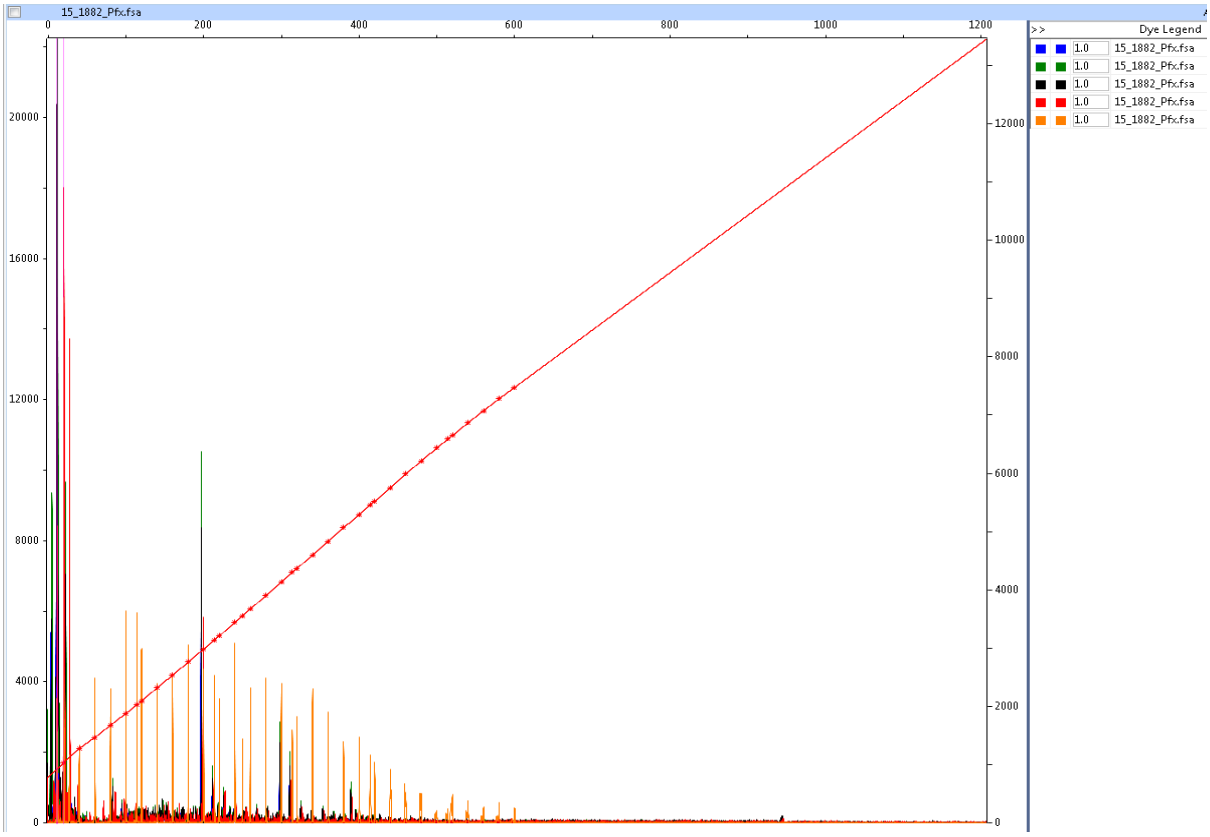


Figure 3-6 Example of a Peakscanner screenshot showing amplification of fragments with correct signal intensity in a sample when Platinum *Pfx* Polymerase proof-reading enzyme mastermix was used

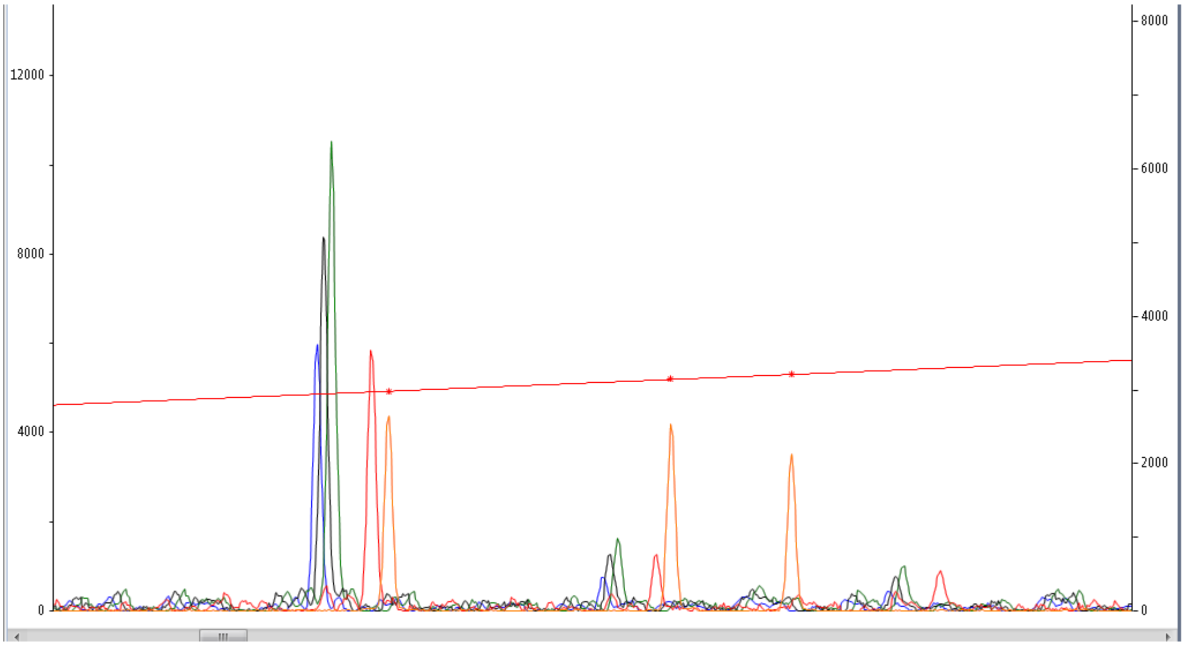


Figure 3-7 Second example of the Peakscanner screenshot showing amplification of fragments from the sample above in the previous figure but a mixture of fragments noticed at a single position leading to difficult interpretation of the data when Platinum *Pfx* Polymerase proof-reading enzyme mastermix was used

3.3.2.4 Recombinant *Taq* Polymerase (Invitrogen, U.K.)

The majority of the samples analysed using this enzyme were amplified with optimum signals apart from a few samples which showed stronger signal intensity (figure 3-8) and double peaks (figure 3-9). The electropherogram for the reference strain, H37Rv, was very clean and the signal intensity was normal (see figure 3-10).The negative control, water, did not show any amplification (see figure 3-11).

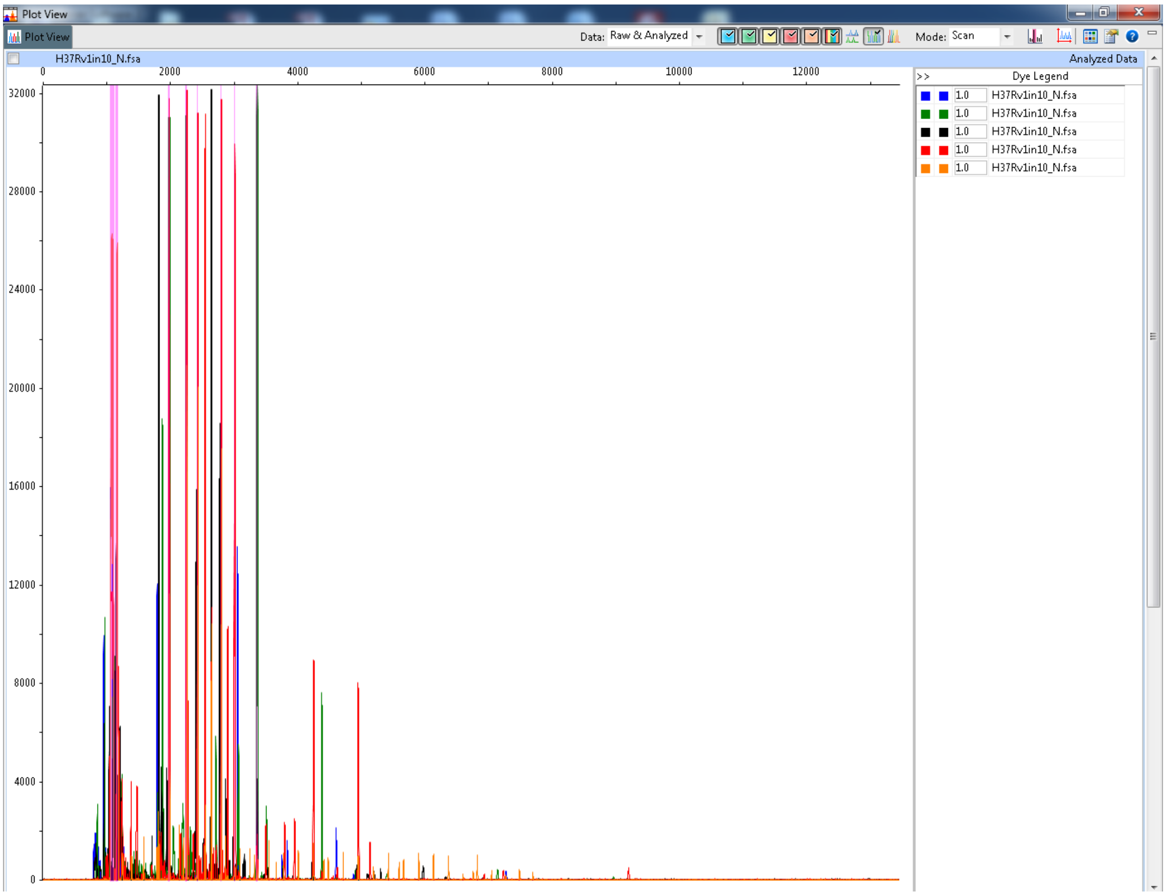


Figure 3-8 Example of a Peakscanner screenshot showing strong raw signal intensity in H37Rv sample using recombinant *Taq* Polymerase reagents

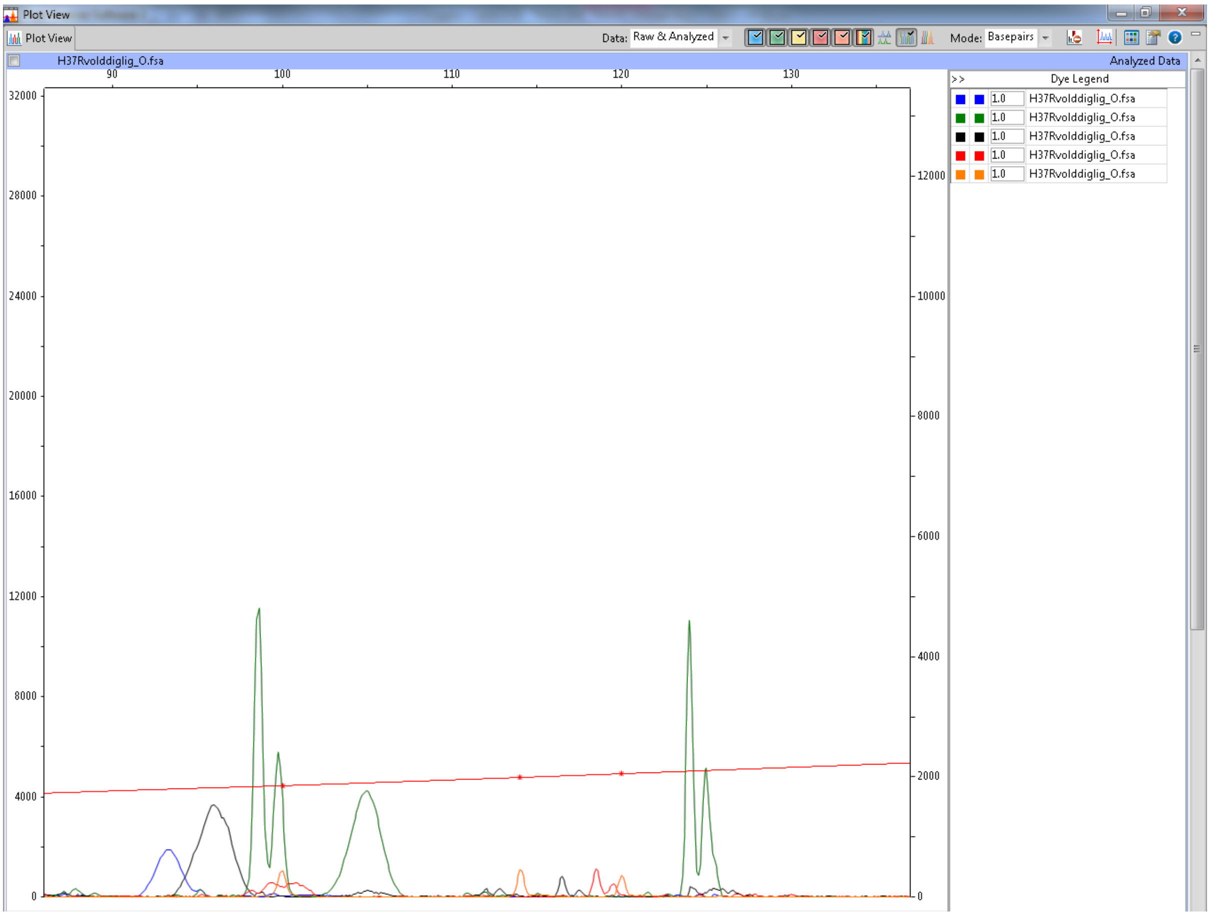


Figure 3-9 Example of a Peakscanner screenshot showing double peaks (green dye- VIC) in H37Rv sample using recombinant *Taq* Polymerase reagents

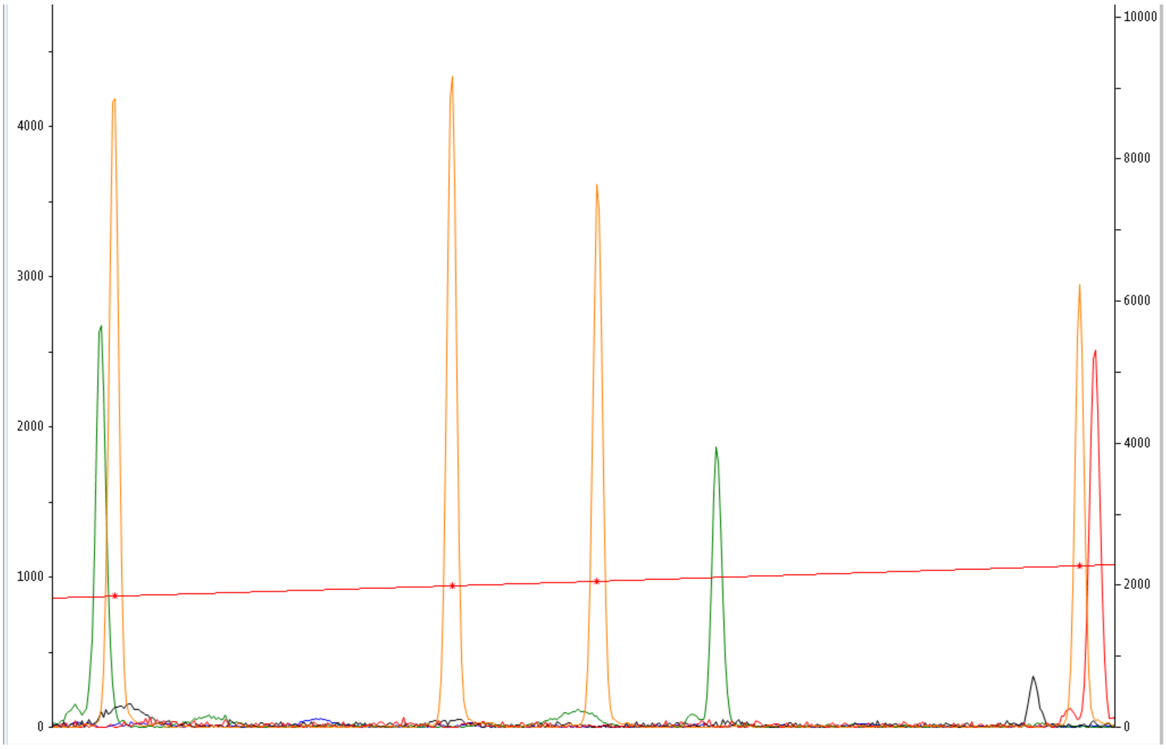


Figure 3-10 Example of a Peakscanner screenshot showing the presence of amplified fragments with single peaks in an H37Rv sample when using recombinant *Taq* Polymerase reagents.

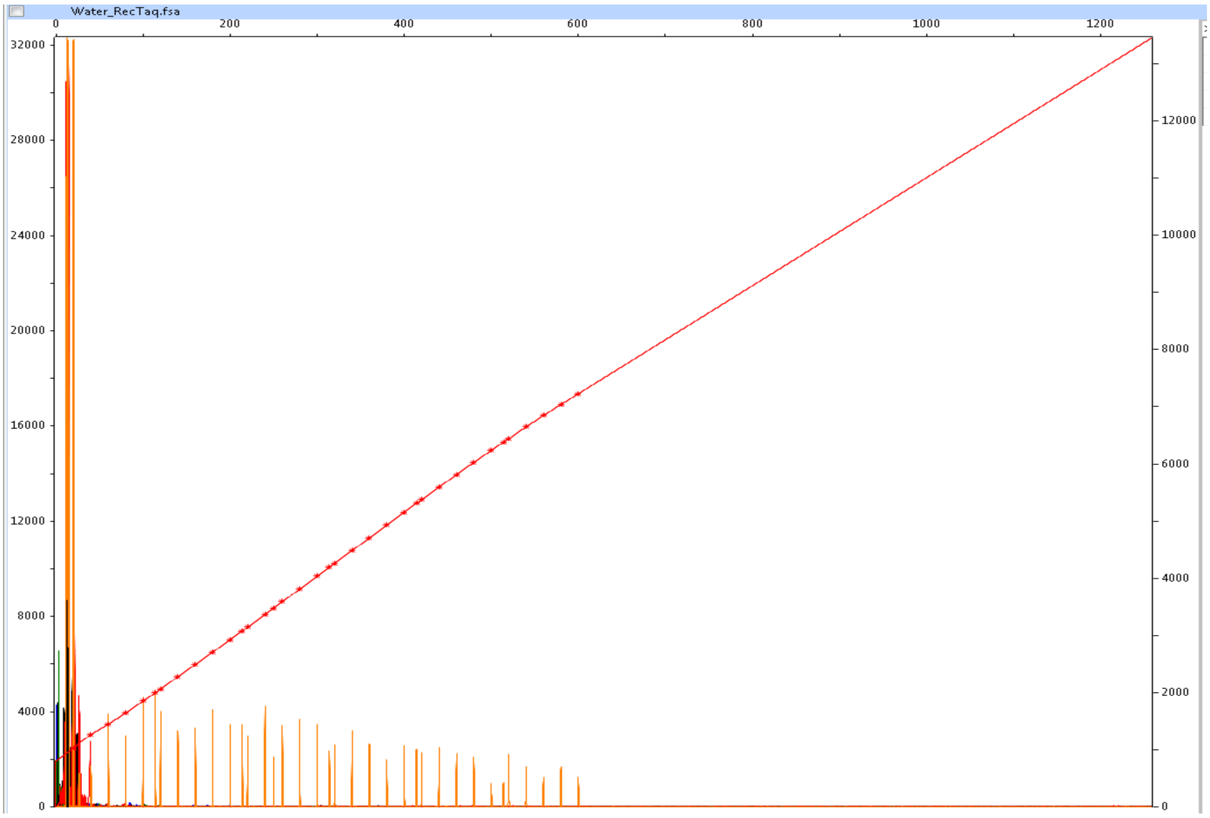


Figure 3-11 Peakscanner screenshot showing the absence of amplification of fragments in the negative sample (water) using recombinant *Taq* Polymerase reagents.

3.3.3 H37Rv and Nepal TB DNA study samples

Changing the PCR conditions significantly increased the efficiency of the IS*6110* FAFLP assay and, together with enzyme optimisation development, the assay was standardised. So, using this technique, H37RV (see figure 3-9) and the 176 Nepal samples were characterised (see appendix 5- raw FAFLP data stored in CD format). It is not possible to include the electropherograms for all the 176 samples due to space constraints but the example below, figure 3-12, shows the amplification of fragments for two samples displaying well-defined fragments without artefacts.

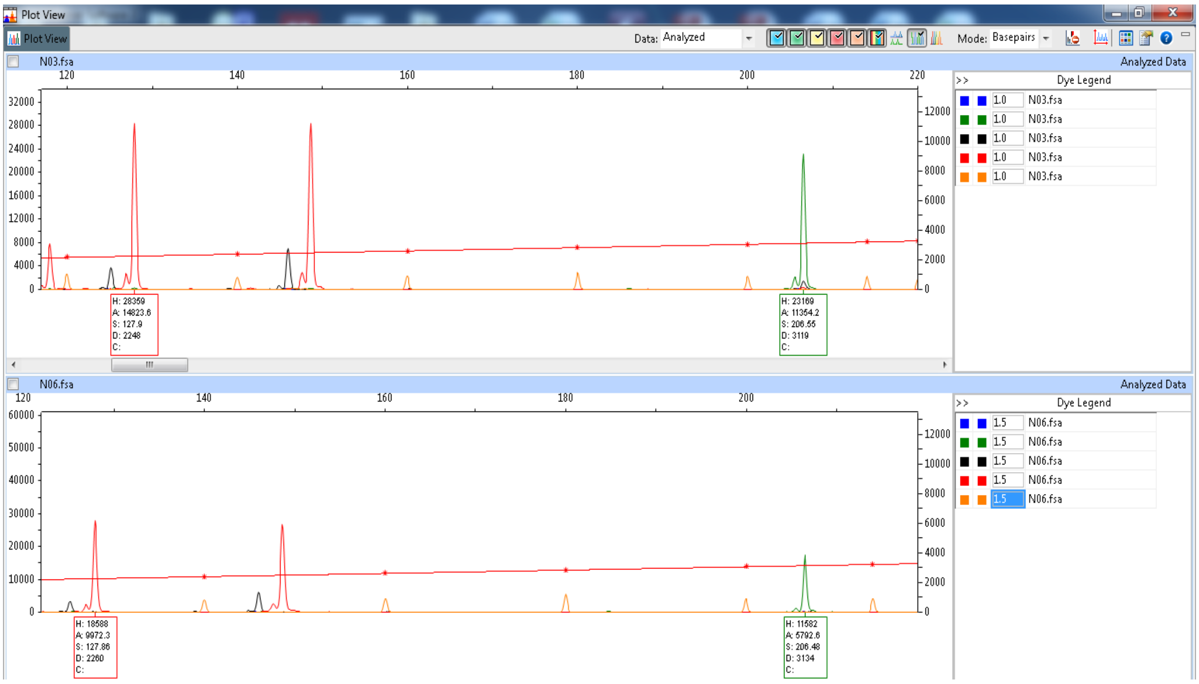


Figure 3-12 Example of an electropherogram showing two Nepal samples, N03 and N06, identical to each other having the same fingerprint with well-defined and easily identifiable coloured fragments.

3.4 Discussion

Whilst testing the samples from already adapter-ligated DNA (178), it was apparent that using poor quality DNA was unsuccessful as there was no amplification of products as this DNA was from the 1999 study. This was overcome by testing fresh DNA samples from H37Rv and Nepal TB strains and the FAFLP procedure was carried out on the samples including restriction enzyme digestion, ligation and PCR. Unaltered PCR conditions and reagents led to the generation of double peaks (figure 3-8), which created difficulty in interpreting the data. It was apparent that the PCR conditions required optimisation, including change of PCR reagents including PCR enzymes. The PCR final extension step was increased from 60°C to 72°C for 15 minutes but otherwise all other conditions remained the same. Of four PCR reagents tested, the fragments were not amplified using Hotstar *Taq* Plus mastermix. With regards to My*Taq* mastermix, the signal intensity of 28000rfu was high and led to the difficulty in interpretation of data. The signal intensity for the third PCR reagent, *Pfx* Polymerase gave optimum signal intensity and also the background noise was low compared to former two reagents mentioned above but there were more than one fragment on one given location which made the analysis difficult. The fourth reagent tested, recombinant *Taq* polymerase mastermix, gave good signal intensity but also produced double bands which indicated that the PCR conditions needed further optimisation. A final extension of 72°C for 60 minutes was carried out instead of 15 minutes as in previous protocols and gave well-defined, easily interpretable fragments and thus the standardisation was complete. This technique was used in chapter 4 for mapping the IS*6110* insertion sites in H37RV genome and in chapter 6 for characterisation of Nepalese TB strains into different TB genetic lineages.

3.5 Summary

In this chapter, the IS*6110* FAFLP PCR procedure was standardised and tested using the adapter ligated DNA products from a global TB collection (178), H37Rv and followed by samples from Nepal. I selected the latter for my thesis to develop the genomic tool box for resource poor settings as Nepalese samples were very poorly characterised when this study was started in 2011.

.

Chapter 4 Mapping of Insertion sites (IS*6110*) in the *M. tuberculosis* H37Rv reference genome and rapid definition of genetic lineages

(This chapter was published in the Journal of Microbiological Methods) (179)

4.1 Introduction

Insertion elements play a significant role in the evolution of genomes of various organisms (69) and can transpose within the genomes thereby altering the position and even the function of the adjacent genes. These transpositional changes have contributed to the evolution of different microorganisms thereby exhibiting a genome plasticity function where the genome is altered to survive in different environmental niches (75,180).

Tracing the movement of insertion sites can be useful in two ways, firstly by acting as genetic markers to define strains of different microorganisms and thus determine the transmission relationship between two strains (181) and secondly by mapping the exact position of IEs whose transposition could alter the phenotype of the strain. Mapping of IS*6110* sites in the Zaragoza *M. tuberculosis* strain that caused a sudden outbreak in Spain in the years between 2001 and 2004, has offered clues of the adaptability and virulence of *M. tuberculosis* (182). The Zaragoza strain belongs to PGG3 and is responsible for widespread outbreaks compared to PGG1 and PGG2 (refer section 1.9.1 for the descriptions about PGGs) in Spain. On further analysis of the mapping results, Millan-Lou et al. have found that there are twelve copies of IS*6110* in this strain and that the IS*6110* insertion in the DR region is found 3bp away from the DR sequence which is different to the compared H37Rv strain. Also, five of the twelve insertion points were at positions already reported by other scientific groups over the years. Most importantly, one of the insertion locations, in the Rv2823c gene was unique and specific to the Zaragoza strain which then led to the development of multiplex PCR assay for the rapid detection of this strain of *M. tuberculosis* (182).

Genomic mapping identified that there are differences between two strains of Beijing family, W and 210. These two strains have seventeen IS*6110* insertion sites, twelve are identical in terms of insertion locations in the genome but five are different. Two unique sites were found in W strain and also one IS*6110* copy was found upstream of the *ctp*D gene wherein IS*6110* was providing a promoter element for the transcription of this gene (98). Alonso et al., found specific IS*6110* insertion sites in the Beijing genotype that helped in the mapping of this family of TB and also showed that the Beijing TB lineage displayed a higher frequency of IS*6110* insertions than other lineages (183). TB genetic lineages, as described in chapter 1 section 1.9, are grouped mainly into two types, either by SNP classification (153) into 3 Principal genetic groups (PGG1, PGG2 and PGG3) or by LSP classification (184) into seven genetic lineages. These genetic families or lineages are important as it has been suggested that there is phenotypic diversity existing between *M. tuberculosis* clinical isolates (165). Also, it has been reported that Beijing/W lineages, with their unique phenotypic traits, have an increased ability to cause disease (185,186). Also, this strain has been shown to be predominantly distributed in Eastern Asia and due to human migration this genotype migrated from Beijing City to the other parts of the world over time. However, a high number of insertions in certain strains does not necessarily suggest pathogenicity but that the molecular clocks of IS*6110* movement is faster in some strains than others, as it has been shown that low copy numbers are also highly pathogenic (107).

As there is a lot of ambiguity in defining a specific TB genetic lineage using the available epidemiological markers, and because the techniques described above are complex in terms of technicality, a rapid definition assay will be helpful in providing information about the TB genetic lineage and possibly then allowing improved management of the disease caused by a particular genetic lineage.

4.1.1 Aims and Objectives

The aim of this chapter is to locate or map the precise genomic IS*6110* insertion sites in the *M. tuberculosis* H37Rv strain using IS*6110* FAFLP PCR methodology and define the TB genetic lineages, H37Rv (PGG3 group) and Beijing (PGG1 group). Rapidly defining TB lineages would help in identification of TB genetic lineages circulating in a given area of the outbreak and thereby indirectly help in the informed management of the disease including control and surveillance.

The following specific objectives are used to achieve the aim mentioned above:

- To carry out IS*6110* FAFLP PCR methodology on H37RV genomic DNA

- To decrease the number of fragments generated by FAFLP to enable sequencing of individual fragments so that it is easy to interpret the insertion points of IS*6110* in the genome.

- To identify the sequence of individual fragments generated and map the insertion site IS*6110* position in the reference sequence of H37Rv.

- To develop and perform lineage specific PCR to define lineages as PGG1-Beijing and PGG3-H37Rv.

4.2 Materials and Methods

H37Rv TB DNA was subjected to IS*6110* FAFLP methodology as described in chapter 3 (chapter 3, section 3.1.3) and the fragments generated by capillary electrophoresis under the run conditions described in chapter 2 (chapter 2, table 2.2), were analysed using PeakScanner software (Thermofisher, UK). To further reduce the number of fragments, new two base selective primers were designed with an addition of another selective base at the 3'end of the existing *Taq*I primers (see table 4-1 below) and subjected to the previously used PCR conditions using IS*6110* reverse primer. One microlitre of the PCR product is used to sequence in both directions using the same primers. The sequence data is then subjected to BLAST analysis to reveal the identity of the fragment thereby allowing the fragment of interest to be mapped at the genome level (see Tables 4-2 and 4-3)

Table 4-1 List of unlabelled two base selective primers

Two base selective *Taq*I primer	**Primer Sequence (Two selective bases at 3' end)**
1) *Taq*I- A**A**	5'- CGATGAGTCCTGACCGA(A**A)**-3'
2) *Taq*I- A**C**	5'- CGATGAGTCCTGACCGA(A**C)**-3'
3) *Taq*I- A**T**	5'- CGATGAGTCCTGACCGA(A**T)**-3'
4) *Taq*I- A**G**	5'- CGATGAGTCCTGACCGA(A**G)**-3'
5) *Taq*I- C**A**	5'- CGATGAGTCCTGACCGA(C**A)**-3'
6) *Taq*I- C**C**	5'- CGATGAGTCCTGACCGA(C**C)**-3'

7) *Taq*I- C**T**	5’- CGATGAGTCCTGACCGA(C**T)**-3’
8) *Taq*I- C**G**	5’- CGATGAGTCCTGACCGA(C**G)**-3’
9) *Taq*I- T**A**	5’- CGATGAGTCCTGACCGA(T**A)**-3’
10) *Taq*I- T**C**	5’- CGATGAGTCCTGACCGA(T**C)**-3’
11) *Taq*I- T**T**	5’- CGATGAGTCCTGACCGA(T**T)**-3’
12) *Taq*I- T**G**	5’- CGATGAGTCCTGACCGA(T**G)**-3’
13) *Taq*I- G**A**	5’- CGATGAGTCCTGACCGA(G**A)**-3’
14) *Taq*I- G**C**	5’- CGATGAGTCCTGACCGA(G**C)**-3’
15) *Taq*I- G**T**	5’- CGATGAGTCCTGACCGA(G**T)**-3’
16) *Taq*I- G**G**	5’- CGATGAGTCCTGACCGA(G**G)**-3’

4.2.1 Rapid definition of TB genetic lineages-PGG3 (H37Rv) and PGG1 (Beijing)

Samples: H37Rv and samples belonging to other lineages already assigned using FAFLP (chapter 5) (N10, N25, N34, N46, N62, N63 N70), were selected (appendix 1) to test the proof of principle. Based on FAFLP, the following lineages were assigned to these samples above (Table 4-2).

Table 4-2 showing the selected samples and their IS*6110* FAFLP derived lineages

Sample ID	**FAFLP derived lineages (PGG)**
N10	CAS – PGG1
N25	Unassigned
N34	T (H37Rv-like)- PGG3
N46	Unassigned
N62	Unassigned
N63	Unassigned
N70	Beijing (PGG1)

Following sequence determination (and therefore specific genomic insertion sites) of lineage specific fragments found only in defined lineage/spoligotype groups by FAFLP, primers were designed to amplify common fragments from the PGG1 (5’-gggctccccagatcaa-3’) and PGG3 (5’-gtgtgccgcgaggtgg-3’) groups and comprised a 200bp product and a 296bp product on a gel respectively, based on the hypothesis that a PCR product of a known size will be generated if the IS element is inserted at the exact same genomic region. In a total volume of 50µl, 1µl of the DNA extracted was added to the reaction containing 1X PCR reaction buffer, 1.5 mM $MgCl_2$, 0.2 mM dNTPs (Invitrogen, UK), 5 µM of the PGG1 or PGG3 group primer, 5 µM of IS*6110* reverse primer as described above and 1U of recombinant *Taq* polymerase (Thermofisher, UK). The following PCR conditions were carried out in a Veriti thermocycler (Thermofisher, UK): 96°C for 15 min followed by 30 cycles of 96°C for

1 min, 60°C for 1 min and 72°C for 30 seconds with a final extension of 72°C for 1 min was carried out before cleaning up the PCR products using AMPure XP magnetic beads. The products were then run on a gel to check the presence of the fragment, 296bp product for the PGG3 group and 200bp product for the PGG1 lineage.

4.3 Results

4.3.1 Prediction of fragments using Seqbuilder (DNASTAR Lasergene version 8, USA) software (*in silico* analysis)

The genomic sequence of H37Rv (accession no: NC_000962.2) was digested *in silico* with *Taq*I enzyme followed by the identification of 59 bases of IS*6110* sequence (CTGACATGACCCCATCCTTTCCAAGAACTGGAGTCTCCGGACATGCCGGGGC GGTTCAG) to indicate the size of the expected product by PCR using primers with a two base extension at the 3' end to reduce the number of fragments generated as shown in Tables 4-3 and 4-4 and Appendix 1. The table shows the presence of sixteen insertion sites corresponding to the presence of sixteen IS*6110* copies in H37Rv as was reported previously by Philipp et al in 1996 (187).

4.3.2 Identification of fragments using Peak Scanner (Thermofisher Scientific, UK) software (*in vitro* analysis)

Sixteen fragments were mapped as seen in columns 1 and 2 in tables 4-3 and 4-4 using their fluorescent dye colours.

4.3.3 Mapping the insertion site IS*6110* position in the genome of *M. tuberculosis.*

All but two, 71.5 Y and 83.8 Y as seen in tables 4-3 and 4-4 in column 1, of the sixteen fragments generated could be identified by this method. Two fragments could not because the flanking genomic data contained in the sequence was too short to analyse as the restriction site was too close to the end of the transposon sequence. Although the two base selective primers are specific and generate a largely specific PCR product for sequencing, a degree of mis-priming does occur. These non-specific extension products are generated due to mis-priming occurring in the PCR. The data generated for the H37Rv genome fragmented using the 4-dye FAFLP PCR in Tables 4-3 and 4-4 and Appendix 1 correlate with the sequencing data and also the *in silico* data generated using DNASTAR Lasergene Seqbuilder as described above.

Table 4-3 Mapping of H37Rv genome using 4-dye FAFLP PCR, DNA sequencing including BLAST results and *in silico* analysis with sequence orientation 5'-3' according to Figure 3-1 Method Schematic

FAFLP fragment colour and size in base pairs*	**Sequence product size**	**Insertion site BLAST result from sequence data**	***In silico* predicted insertion site**
99.1 G	(102)	Transcriptional regulator for LacI family protein	Transcriptional regulator for LacI family protein

81.0B	(87)	Too short to analyse	Cutinase1
173.6R	(175)	Conserved membrane protein	Conserved membrane protein
736.6R	(772)	IS*6110*	IS*6110*

*** G=Green, B= Blue, R=Red, Y=Yellow**

Table 4-4 Mapping of H37Rv genome using 4-dye FAFLP PCR, DNA sequencing including BLAST results and *in silico* analysis with sequence orientation 3'-5' according to Figure 3-1 Method Schematic.

FAFLP fragment colour and size in base pairs*	**Sequence product size/*in silico* predicted size in base pairs**	**Insertion site BLAST result from sequence data**	***In silico* predicted gene insertion site**
206.4Y	(201)	Peroxidase BpoA	Peroxidase BpoA
140.4R	(141)	Molybdenum cofactor MoaC	Molybdenum cofactor MoaC
560.4B	(561)	PPE-family protein	PPE-family protein

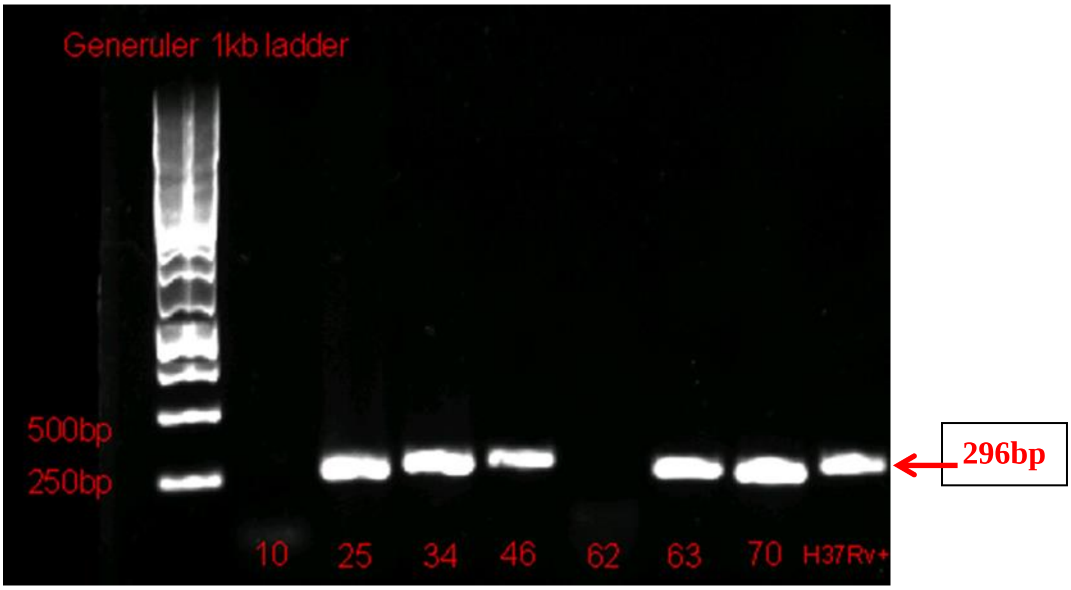
191.7R	(196)	Conserved hypothetical protein	Conserved hypothetical protein
71.5Y	(58)	Too short to analyse	Cutinase1
160.3Y	(164)	LytB related protein	LytB related protein
83.8Y	(58)	Too short to analyse	
Not seen in FAFLP data	(289)	Putative transposase for IS*986*	Putative transposase for IS*986?*
Not seen in FAFLP data	(214)	Integrase	Integrase?
329.2B	(338)	Glycerolphosphodiesterase	Glycerolphosphodiesterase?
124.9G	(127)	Transposase	Transposase
360.0R	(360)	Oxidoreductase	Oxidoreductase

*** G=Green, B= Blue, R=Red, Y=Yellow**

4.3.4 Rapid definition of TB genetic lineages- PGG3 and PGG1.

For the seven samples tested as proof of principle, five belonged to the PGG3 lineage: H37Rv, N25, N34, N46, N63, one belonged to the PGG1 lineage, N70 according to FAFLP data, one belonged to the CAS lineage, N10 and one (N62) did not give any result (see figure. 4-1). N70 reacting to PGG3 primer shows there might be fragments shared by different PGG groups and needs to be looked into detail.

A.



B.

Figure 4-1 Agarose Gel Electrophoresis showing *M.tuberculosis* strains , N10, N25, N34, N46, N62, N63, N70, H37Rv, amplified using PGG3 (A) and PGG1 (B) specific primers, amplifying 296 bp and 200 bp products respectively

4.4 Discussion

Although a large number of studies have been published based on the fragment patterns generated by IS*6110* RFLP, little or no data are available on the distribution of IS*6110* genomic insertion sites in *Mycobacterium tuberculosis* complex strains. This will have an impact on the molecular epidemiological studies of *M. tuberculosis*. Knowledge of specific sites may lead the way to develop rapid techniques to identify specific lineages which could be validated in future using WGS technology.

This technique exploits the use of selective bases at the 3’ end of primer sequences to reduce the fragment numbers generated during FAFLP PCR amplification. Also, use of differentially labelled fluorescent primers aids the identification of fragments. Using selective primers reduces the number of fragments amplified, which can then be sequenced and the position of the insertion site in the genome identified by BLAST analysis of the sequence. The BLAST results show that, except for the three fragments (81B, 71.5Y and 83.8Y) that were too short to analyse, the predicted and actual insertion sequence data correlate with each other and the FAFLP fragment size and colour. As every fragment generated always contained 75 bp of transposon/adaptor sequence, any fragments below 100bp were too small to identify flanking genomic sequence by this method. Sequencing data usually starts around 25 bp past the sequencing primer which effectively cuts off 100bp in total. Apart from the fragments too small to analyse (this could be remedied by choosing an alternative enzyme for FAFLP), there were three fragments that were not predicted in silico. These were an IS986 transposase (corresponding to 289bp fragment in table 4-3), integrase (214bp in table 4-3) and glycerol phosphodiesterase (329.2B in table 4-3), involved in the glycerol metabolic process. That another transposon was identified strongly suggests that a further transposition has occurred following the many cultural iterations of H37Rv since whole genome sequencing as seen in the case of IS986 transposase. If this was the case then these fragments would not be predicted from the genome sequence available. The predicted insertion sequences seen offer interesting avenues for further study, including characterisation of the surface genes disrupted by IS*6110* including PPE and those affecting metabolism. The characterisation of insertion sites of this and other transposons is important for clonal organisms such as those belonging to the MTBC.

In 2011 Alonso et al., showed the importance of mapping insertion sites by demonstrating that the IS*6110* in the Beijing strains of MTBC can up-regulate downstream genes via an outward-directed promoter in its 3'end (Alonso et al. 2011). Further mapping of the specific insertion sites will generate information on the nature of the gene disruption, whether the insertion has a detrimental effect, for example disrupts the proposed reading frame, or has a potentially beneficial effect through up regulation. The next steps for this work after it has been shown to map successfully the IS*6110* insertion sites in H37Rv is to apply this knowledge to design a rapid tool for the definition of lineages.

The principle of rapid detection of lineages was successful using the lineage specific fragments targeting the common fragments as described above (see figure 4.1). This assay needs a simple PCR thermocycler, reagents and facilities to run and visualise the gel and the ability to interpret the results. Direct detection from sputum would be a useful next step but sensitivity is likely to be low. Detection of lineages by the LSP or the SNP methods are time consuming as they need good quality DNA after bacterial DNA extraction and a bioinformatics approach to deduce the TB families.

They also need expensive instruments like the sequencers and extra manpower to accomplish the definition of lineages. Although an alternative approach would now be made if starting the project, the FAFLP approach was the only approach available at the time and as such it still forms the basis of genomic tool box development for the rapid assignation of TB genetic lineages based on IS*6110*.

4.5 Summary

The data presented shows that the IS*6110* FAFLP PCR technique, though relatively simple, is robust and effective in mapping IS*6110* insertion in H37Rv that could be applied to any bacterial species with similar repeated elements. Also in this chapter, rapid definition of TB genetic lineages using newly designed primers were carried out using PCR without the use of digestion and ligation, for the rapid detection of both H37Rv-like (T-group) and Beijing lineages to initiate the development of a suite of PCRs based on detection of lineage specific common IS*6110* insertion points as a typing tool for resource poor settings.

Chapter 5 Classification of Nepalese TB clinical isolates into different TB genetic lineages

(This chapter was published in the open access journal of Clinical Microbiology and Infectious Diseases (188)*.*

5.1 Introduction

Nepal is geographically interlocked between China and India, which together account for approximately a third of annual global new cases (11% and 24%, respectively) (WHO 2014) and is ranked 43rd in the world in terms of age-adjusted death rate, which is 27.80 / 100000 of the population. During the year ending of 2014, there were 5506 deaths in Nepal due to tuberculosis that accounts to 3.5% of total deaths in Nepal. TB cases and deaths are more prevalent in men than women. In 2006 Nepal's National Tuberculosis Programme (NTP) developed the STOP TB strategy to fight against tuberculosis, in line with WHO. The NTP is fully integrated with the national healthcare policy of the government of Nepal. The Millennium Development Goal (MDG), another initiative by WHO to tackle TB, was found to be effective in halting and reversing global TB prevalence. The prevalence rate (all per 100,000 population) has gone down from 348 (162-602) in 1990 to 211 (99-365) in 2015. The death rate (all per 100,000 population) has also reduced from 52 (32-70) in 1990 to 19 (13-25) in 2015. Due to all these efforts, Nepal is no longer in the list of high burden TB countries. TB programs were able to save 32,973 lives in 2014 but despite this progress, there were still 978 deaths. Overall treatment success rates of drug susceptible TB was 91.5% with 0.92 % failure rates, 2.17% failed to follow-up and 2.7% death rates. The proportion of MDR-TB was 2.2% among new TB cases and 15.4 % among retreatment cases based on a survey carried out in 2011-2012. New surveillance studies have not been carried out recently. There were a total of 22 deaths among MDR cases and 3 deaths in XDR reported in 2014/15. For MDR the drug resistance pattern showed higher levels of resistance in fluoroquinolones and 8% of those MDR patients further developed XDR. Sputum microscopy is still the gold standard method for the diagnosis of TB in Nepal. Currently there are 581 microscopy centres providing this smear microscopy service throughout the country. As a high proportion of rifampicin resistant strains are resistant to Isoniazid, detection of rifampicin resistance can be used as a marker for MDR-TB as described in chapter 1. Between 2011 and 2012, GeneXpert MTB/RIF assays were introduced in 3 centres and now there are 26 centres nationally (173). The STOP TB program has reduced the number infected from 29 million people, as estimated by WHO in 2014, to 15 million people according to the recent report by NTC. Despite the efforts to control the spread of TB, there is still a long way to go to achieve the goals of STOP TB program as 40,000 people are infected every year, with 20,000 new primary cases and 5000-7000 deaths each year from tuberculosis in Nepal.

It is important to understand the molecular diversity of the *M. tuberculosis* population in Nepal as it has been reported recently that there is a high similarity between TB strains in Nepal and in Northern India, with which Nepal shares an open border policy of human migration (189). This might indicate that the TB lineages that were found in India, especially the CAS lineage, have successfully established in Nepal due to the human migration from India into Nepal (190), enabling the TB lineages to be successful in any geographical location (189).

Very limited data are available on the characterisation of *Mycobacterium tuberculosis* strains and genotypes circulating in Nepal. A recent study of 261 Nepalese isolates found drug resistance in 12.8% of *M. tuberculosis* strains that were from new untreated cases, with the most frequent lineages reported as CAS/Delhi (40.6%), East Asian (including Beijing) (32.2%), Euro-American (15.7%) and Indo-oceanic (11.5%)(191). To gain further insight into the characteristics and diversity of mycobacteria in Nepal, this study aimed to categorise isolates for the first time using IS*6110* FAFLP PCR and to assign them to different genetic lineages.

5.1.1 Aims and Objectives

The main aim of this chapter is to classify the Nepal samples into different TB lineages to understand the genetic distribution of TB lineages in Nepal.

The following objectives were used to achieve this aim:

- To perform IS*6110* FAFLP PCR on Nepal samples
- To characterise the samples into different TB lineages

5.1 Materials and Methods

5.1.1 Strains

(Refer General methods section 2.1

Sputum samples from 176 consecutive new TB patients were collected between 2007 and 2008 and cultured alongside routine diagnostic testing from two Nepalese tuberculosis reference centres located in the Kathmandu valley: the National Tuberculosis Centre (NTC) and the German Nepal Tuberculosis Project (GENETUP). The patient population represented local and referred cases from across Nepal. Bacterial genomic DNA from isolated strains was extracted by the Cetyltrimethylammonium Bromide (CTAB) method.

5.1.2 IS*6110* FAFLP PCR, Fragment Sizing & Analysis

(Refer to the General methods section 2.2 for IS*6110* FAFLP methodology and fragment sizing.). The four-dye FAFLP data collected from the different profiles were recorded and compared with a well characterised reference collection of *M. tuberculosis* isolates (178) using BioNumerics software v6.1 (Applied Maths Inc., Belgium). These data were then used to build a dendrogram using the Dice coefficient of similarities to compare the similarity matrix and Unweighted Pair Group Method with Arithmetic Mean (UPGMA) derived cluster analysis with cophenetic correlation for the branch quality. Dice similarity coefficient and UPGMA clustering method was chosen because of the high correlation in terms of their similarity obtained (in this case, genetic similarity) between the samples to form a group or cluster compared to the simple matching coefficient.

5.2 Results

5.2.1 Analysis of Data using BioNumerics software v6.1

Of the 176 DNA extracts from isolates analysed, the majority of the samples (97, 55.4%) belonged to either the spoligotype-defined CAS lineage (64, 36.6%) or the Beijing lineage (33, 18.8%) grouping under PGG1 and the rest of the samples group under either PGG2 (1.7% S, 3.97% X, 7.95% Haarlem and 2.27% LAM, 2.27% T-Uganda) or PGG3 (2.27% of T) (Table 5-1). Forty three samples (24.4%) were “unassigned”. Common fragments specific to different TB lineages seen were exactly the same as the earlier published report by Thorne et al, (2011) except for an additional fragment, 78.4 G, for the CAS lineage (135). These fragments specify different insertion points where IS*6110* has inserted or transposed. A dendrogram was generated using only the IS*6110* FAFLP data confirming again the above mentioned lineages in relation to the PGGs (Fig. 5-1).

Table 5-1 Common fragments identified using IS*6110* FAFLP PCR in TB genetic lineages of the 176 bacterial DNA isolates in Nepal

PGG-spoligotype/sub-lineage	**Common fragment sizes**	**No. Of Nepal strains (Total=176)**
PGG1-CAS	78.4G, 92.0B, 117.9R, 206.2G, 275.1R	64
PGG1-Beijing	101.7B, 102.5Y, 139.1R, 180.7Y, 254.8G, 332.4R, 353.5B	33
PGG2 –Haarlem	87.0Y, 89.4G, 148.7B, 300.2R (H/X), 445.7Y	14
PGG2-LAM	71.5Y, 105.2R, 116.1R	4
PGG2-S	88.0G, 112.9R, 217.2R, 445.3G	3

PGG2-X	83.8Y, 300.2R (H/X)	7
PGG2-T Uganda	88.9Y, 119.5G, 122.9Y, 228.4Y, 266.8R	4
PGG3- T	81.3R, 192.4R, 360R	4
Ungrouped		43

(PGG represent Principal Genetic groups according to Sreevatsan et al, spoligotypes follow spolDB4 classification and spoligotypes derived sub-lineages are grouped following Gagneux's classification (153,158)).

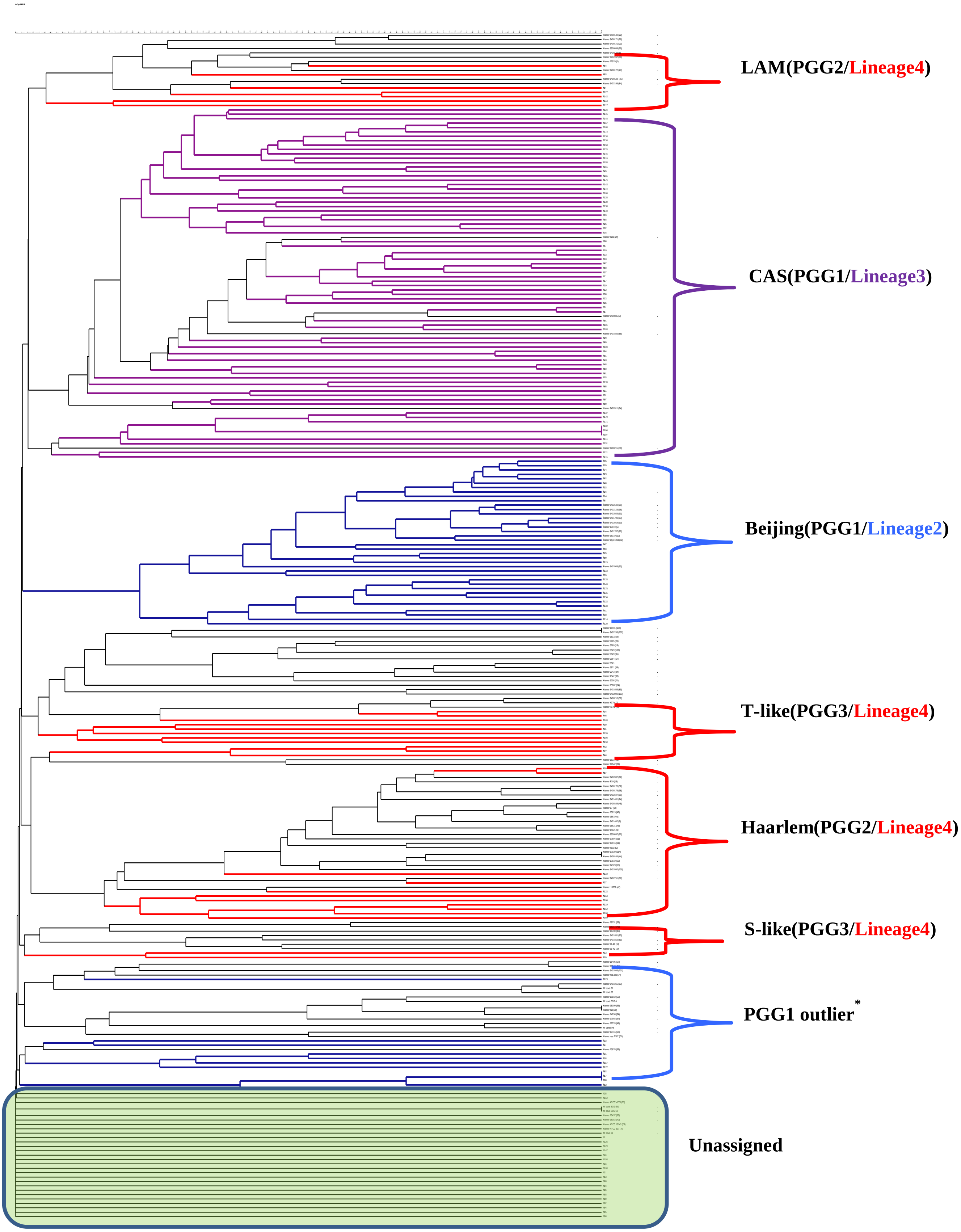


Figure 5-1 UPGMA derived dendrogram showing the predominant genetic lineages/spoligotypes of 176 Nepalese *Mycobacterium tuberculosis* isolates.

Coloured branches represent Nepalese samples following Gagneux's global phylogeography of MTBC (158) and black branches represent the in-house global *Mycobacterium tuberculosis* collection. PGGs are numbered following Sreevatsan's classification (153). PGG1 outliers share one IS*6110* copy with the PGG1 group. Those which contained one IS*6110* copy but could not be assigned to any group (unassigned group in the figure) are shown within the green box.

5.3 Discussion

Thorne et al. showed that IS*6110* FAFLP PCR can be used to delineate the phylogeny of MTBC as shared common fragments can identify the different lineages in a geographical location by comparison with a reference database collection(135). As very limited lineage information is available from strains in Nepal (up until 2012 there was none), the IS*6110* method (published recently) was applied to map the IS*6110* sites in H37Rv (179) and also carried out rpoB sequencing to further characterize strains from this important region.

Fifty five percent of the 176 Nepalese strains analysed belong to the CAS (36.6%) and Beijing (18.8%) modern genetic spoligotypes (PGG1). The remaining 24.4 % of the samples belong to the PGG2 and PGG3 groups (Haarlem, LAM, S, X, T-Uganda and T). However, a limitation of this technique is its difficulty to characterise the samples with less than 4-5 copies of IS*6110* as seen in the unassigned group (24.4%) in figure 5-1, which can be overcome by the use of other typing techniques like MIRU-VNTR(146). The geographical position of Nepal is likely to have influenced this distribution of lineages, with a mixture of predominantly Beijing lineage from the North of the Himalayas and the CAS lineage from the south (190). The fairly high percentage of mainly European lineages (Haarlem, LAM and T, 12.5%) indicates that there has also been mixing of the different lineages over an extended time and that European travellers/migrants to South East Asia and Nepal may have transmitted European strains to the local population, probably due to the rise in globalisation leading to increased human migration between countries as reviewed by Soto (192). The IS*6110* FAFLP data from our study supports the hypothesis that the geographic location of Nepal is the key for the circulation of PGG1 TB lineages, CAS and Beijing, which were predominant in India and China respectively. This shows that the human migration from India and China into Nepal has helped *M. tuberculosis* lineages, especially Beijing and CAS, establish in this country. A large number of people flow between these regions due to various reasons, for example cheaper medical treatment facilities, work, study, trade, pilgrimage and cultural visits. It has been shown that there is a high prevalence of CAS lineages from North India circulating in Nepal due to the migration of Indian population from this region into Nepal for the purposes mentioned above (163,189,190). Likewise, the other predominant lineage, Beijing, found widely distributed in China and Tibet, is seen in Nepal (164,193,194). This simple and informative PCR-based molecular epidemiological technique might prove useful for the study of outbreaks of the disease and importantly also to detect cross-contamination between different strains or isolates in resource poor settings, aiding cluster investigation and possibly informing outbreak management.

5.4 Summary

To summarise, in this chapter I utilised the IS*6110* FAFLP methodology successfully on DNA from TB samples from a resource poor setting, Nepal, and characterised the samples into different TB lineages or genotypes. This method also showed that the majority of the Nepal samples belonged to the PGG1 groups particularly, CAS and the Beijing.

Chapter 6 Rifampicin Resistance status in Nepalese TB isolates from clinical samples

(This chapter was published in the open access journal of Clinical Microbiology and Infectious Diseases (188). *Also, this methodology was employed by another PhD student from Lahore University (Pakistan) supervised by me in a recent paper* (195) *whilst determining the rifampicin status from clinical samples in Pakistan).*

6.1 Introduction

6.1.1 Drug Resistant TB in Nepal (adapted from NTC, 2015)

A recent Drug Resistance Survey in Nepal conducted between 2011 and 2012, has shown that DR-TB levels are increasing with nearly 9.3% of new patient resistant to at least one drug. The issue of increasing proportion of resistance to fluoroquinolones (26.4%) is a major public health concern in Nepal. Among MDR cases, 8% of the cases were found to be XDR due to the fluoroquinolones resistance. So to prevent the increased death rate due to XDR-TB in Nepal, the government is performing DST for second line drugs of all MDR-TB cases at the start of treatment (173). Early case finding is important to decrease human suffering, duration of disease and / or control of DR-TB, decrease financial burden and improve treatment outcome (23). As part of the STOP TB strategy in 2010, Nepal is successfully managing to control TB due to the presence of 581 microscopic centres, 2 solid culture laboratories including capacity for first line (FLD) drug susceptibility testing (DST), first line probe assay (LPA), one liquid culture and one second line drug (SLD) DST facility and 26 GeneXpert centres spread all over this mountainous country. This is in addition to 2 other culture and DST facilities (NTC and GENETUP) functioning in Kathmandu valley.

During the course of this research work the study of 261 Nepalese isolates by Malla et al., found that around 8.04% of the MDR-TB cases were new untreated cases. (191). This study reports that the two major TB genotypes circulating in Nepal are CAS and Beijing. According to the NTP’s annual report published in 2015, the national rate of MDR is low in new cases (2.2%)

and high in retreatment cases (15.4%). Since rifampicin resistance has been widely accepted as a ‘surrogate’ marker for MDR (53,196), it is useful to investigate the MDR-TB status in this set of samples and compare them to the results of Malla et al., as they are new untreated TB cases and, moreover, the phenotypic drug susceptibility test results were unavailable for these samples. It is also been reported that the predominant strains circulating in India and China, CAS and Beijing, are mostly MDR-TB and the same strains are also found in Nepal (164,189,190,193,194). It will be valuable to investigate if this finding is supported with this set of samples using the Rifampicin Resistance detection assay targeting the RRDR region of the *rpoB* gene.

6.1.2 Aims and Objectives

The aim of this chapter is to determine the status of rifampicin resistance as an indicator of Drug Resistance in the 176 bacterial DNA samples from Nepal.

6.2 Materials and Methods

6.2.1 *rpo*B Analysis

The 81bp Rifampicin Resistant Determining Region (RRDR) of the *rpo*B gene in the *M. tuberculosis* genome of all strains was sequenced using published primers (Arnold et al. 2005) and analysed in BIOEDIT software using ClustalW alignment parameters (see appendix 3). The PCR was carried out in a total volume of 50µl where 1ng of the DNA was added to the reaction containing 1xPCR reaction buffer, 1.5mM $MgCl_2$, 0.2mM dNTPs (Thermofisher, UK), 20µM each of both *rpo*B-RRDRforward (5’-

CGATCACACCGCAGACGTTGA) and reverse primers (5'-GGCACGCTCACGTGACAGACC) and 5U recombinant Taq polymerase (Thermofisher, UK). The following PCR conditions were carried out using a Veriti thermocycler (Applied Biosystems, UK): 94˚C for 2 min followed by 35 cycles of 94˚C for 30 sec, 60˚C for 30 sec and 72˚C for 1 min. Finally, an extension of 72˚C for 10 min was performed before cleaning the products using AmpureXP magnetic beads (Beckman Coulter, UK) and sequenced using the forward primer, *rpo*B-RRDR forward.

6.3 Results

Of 176 DNA extracts analysed for *rpo*B mutations, seven samples (3.9%) had a single non-synonymous base change which would likely confer resistance (see table 6-1 and appendix 3). Six of these seven samples also showed a second base mutation in a codon triplet whereas sample N70 showed a first base mutation. There were no silent mutations observed in the RRDR region of any of these samples.

Table 6-1 List of mutations seen in *rpo*B Rifampicin Resistance- Determining Region (RRDR) of rifampicin resistant *M. tuberculosis* isolates from Nepal.

Sample	**Mutated locus**	**Nucleotide modification**	**Amino acid modification**
1. N70	516	GAC > TAC	(Asp>Tyr)
2. N10	522	TCG > TTG	(Ser>Leu)

3. N25	526	CAC > CTC	(His>Leu)
4. N63	526	CAC > CGC	(His>Arg)
5. N34	531	TCG > TGG	(Ser>Trp)
6. N46	531	TCG > TTG	(Ser>Leu)
7. N62	531	TCG > TTG	(Ser>Leu)

6.4 Discussion

Inferred Rifampicin resistance status for all the Nepal samples was determined using *rpo*B analysis. According to a recent study (191), fifty strains had any drug resistance and sixteen (6.1%) out of 261 isolates were MDR. Among the fifty any drug resistant strains, 29 cases were previously treated and twenty-one were new untreated cases (8.04% of 261 total strains and 12.8% of 164 new untreated cases). In this study, MDR-TB was tested by using rifampicin as the resistance marker and 7 isolates out of 176 were found (3.9%) from new untreated cases possessing drug resistance genotypes. However, their reported MDR percentage was based on the total number of isolates, of which 37.2% of isolates were from previously treated cases, which may enhance a probability of drug resistance development compared to untreated cases. Our results indicate that the prevalence of Rifampicin (RIF) resistant TB (surrogate marker for MDR) was higher than the nationally reported 2.2% MDR in new untreated cases. Further, our results are concordant with a recently conducted study by Creswell et al., where they have shown that the genotypic rifampicin resistance in newly diagnosed TB patients to be 3.3% in Nepal (198). In this study, the most common mutation site in the RRDR is at codon 531 concordant with the global data and parallels the findings of earlier studies (190,197). It also demonstrated the successful use of a surrogate rifampicin marker in analysis of MDR in *Mycobacterium tuberculosis* strains isolated from newly diagnosed primary TB patients originating from different regions of Nepal.

The major circulating genotypes in Nepal according to our study have been shown to be CAS and Beijing in chapter 5 table 5-1 which was concordant with Malla’s study (191) and Sharma’s study (189). Sharma et al. reported that the major circulating genotype in northern India was CAS and they were predominantly MDR (189). As Nepal shares an open border with India in the south, this had contributed to the spread of the CAS lineage into Nepal and thereby the MDR-TB (189,190). Again the same scenario is true for the Beijing lineages from China contributing to the MDR burden in Nepal (164,189,190,193). Despite the unavailability of the drug susceptibility tests, this rpoB assay was able to infer the rifampicin resistance status for this set of samples.

The combination of the rapid lineage specific PCR assay as discussed in chapter 4 along with the RIF resistant PCR assay will be more useful in resource limited setting as there is a reduced requirement for highly specialised equipment/infrastructure and staff for these molecular tests.

6.5 Summary

In this chapter by utilising the rpoB PCR assay targeting the 81bp RRDR region, seven samples (3.9%) of the 176 Nepal samples tested were found to be rifampicin resistant and is concordant with the other studies as described above.

Chapter 7 Final Discussion

7.1 Introduction

Control of tuberculosis in high burden countries is of paramount importance for a global TB prevention strategy to be successful. India and China, along with Indonesia, account for 45% of the new TB cases according to the global tuberculosis report by WHO (12). Despite recent advances in drug development and rapid diagnostic assays, the majority of TB diagnostic centres are still dependent on clinical diagnosis and interpretation of bacterial cultures to formulate the drug regimen and treat TB infection (12).

7.2 Discussion of findings in this study

The hypothesis of this research work is to substantiate the analysis of the position and copy number of IS*6110*, a bacterial transposon, as a genomic tool to characterise the TB genetic lineages from Nepal, one of the low and middle income countries. TB strain typing can support molecular epidemiological investigations in controlling onward transmission and outbreaks, by identifying patients who are linked in the same chain of TB transmission (PHE, 2014).

7.2.1 Optimisation and Development of Insertion Element IS*6110* Fluorescent Amplified Fragment Length Polymorphism (FAFLP) PCR

The main aim of this thesis was to develop a molecular tool box using IS*6110* FAFLP PCR that can characterise *Mycobacterium tuberculosis* complex strains genotypically and be used to understand the distribution and evolutionary relationships between different TB genetic lineages in resource poor settings in any geographical location. Various practical issues/artefacts were addressed by using a recombinant *Taq* polymerase and by increasing the extension temperature from 60°C to 72°C and the extension time from 15 minutes to 1 hour, as shown in chapter 3. The increase in extension temperature reduced mis-priming of the polymerase, producing cleaner, more discrete products for more effective downstream analysis.

7.2.2 Mapping of Insertion sites (IS*6110*) in the *M. tuberculosis* H37Rv reference genome and rapid definition of genetic lineages (published work- (179))

To test the developed assay the reference TB genome H37Rv, was mapped *in silico* and the predicted fragment sizes compared with laboratory results. In this study laboratory results showed that sixteen IS*6110* insertion sites were found in H37Rv (chapter 4 figure) using IS*6110* FAFLP PCR, which correlates with the study by Philipp et al. in 1996 (187), who showed that H37Rv has sixteen IS*6110* insertion sites by pulse field electrophoresis. Here, all sixteen fragments were identified using a fluorescently labelled dye as shown in table 4-2 in chapter 4, together with the position of *in silico* predicted insertion sites.

The IS*6110* insertions have predilection to insert into mainly the surface proteins PPE (560.4B in table 4-2), cutinase1 (71.5Y in table 4-2) and those affecting glycerol metabolism such as IS*986* and integrase (table 4-2) which was recently shown in Roychowdhury's study, where it is shown that the majority of the transpositions of IS*6110* occur in the surface proteins especially the PE/ PPE family of proteins (74). IS*6110* has been shown by Alonso et al. to upregulate downstream genes with the help of the promoter at its 3’end (199). One of the important properties of IS*6110* is that it carries a promoter element upstream of the coding region, which has the potential to up-regulate expression of downstream genes (74,199,200). In a recent study, 178 unique genes belonging to mainly PPEs / transposases and oxidoreductases, were found to be carrying the IS*6110* promoter element upstream of the coding regions (74). This study also found one instance of IS*6110* insertion upstream of the *trp*D gene (ATP-binding cassette transporter), a metabolic enzyme essential for the bacterial survival in activated macrophages whilst colonising the lungs during infection (74,201). Further analysis revealed insertions in MDR-TB isolates, 75bp upstream of the *pho*P region (a transcriptional regulation factor), suggested to be necessary for *M. tuberculosis* virulence by Soto et al. in that the IS*6110* insertion upregulates the over expression of *phoP* gene in *M. bovis* B strains (200). IS*6110* insertions have been found upstream of early secretory antigenic target protein 6 (ESAT-6) that has been linked with mycobacterial virulence (202). It has been suggested that ESAT-6 interacts with the host protein Beta 2 microglobulin (β2M), a serum protein found associated with Major Histocompatibility Complex-1 (MHC-1), thereby inhibiting the expression of β2M- MHC-1 and down regulating the antigen presentation function of class I MHCs (202,203). The identification and confirmation of IS*6110* insertions *in silico* into different genes like PPE, IS*986*, cutinase1, integrase, oxidoreductase and glycerolphosphodiesterase as shown in table 4-2 in chapter 4, indicates that IS*6110* transposition might play a role in the pathogenesis of TB as suggested by other groups and the work presented in this thesis supports that theory.

In addition to the insertions at different genes, common IS*6110* insertion points or fragments occur between lineages but some of them are unique to specific lineages as shown in chapter 5 and reported earlier by Thorne et al. (135). In this study, it has been shown that there are common IS*6110* insertion points / fragments between isolates in one lineage, but not shared with other lineages (see table 5-1), for example Lineage 3, PGG1-CAS (78.4 G, 92.0B, 117.9R, 206.2G, 275.1R) does not share the same fragments as Lineage 4, PGG2-Haarlem (87.0Y, 89.4G, 148.7B, 300.2R, 445.7Y) or PGG3-T (81.3R, 192.4R, 360.0R,) This work was published in 2016 (188) and is concordant with the earlier study by Thorne et al. In the major study by Roychowdhury et al., a computational pipeline was developed to analyse the insertion sites of 1377 whole genome sequenced *M. tuberculosis* isolates, representing the 7 major global lineages, from all publically available datasets. Using density distribution studies, they found that different lineages have different copy numbers and insertion points and these insertions are unique to particular lineages (74). This property was observed in this study prior to Roychowdhury et al., as shown in chapter 5, where common fragments occur between lineages, together with some unique fragments within lineages. Part of the work in this thesis was cited by Roychowdhury et al. as being the first to identify this. In the East Asian- (Lineage 2, PGG1-Beijing) and Indian-East African (Lineage 3, PGG1- CAS) lineages, there is a high level of conservation in insertion sites as well as noticeable differences. Lineage 2, PGG1-Beijing isolates have extra IS*6110* insertions in regions like Rv0001-Rv0002, Rv1371 and *idsB*, involved in biosynthesis of membrane ether-linked lipids (199). In Lineage 3 isolates, there are unique intergenic regions mapping to regions such as Rv0395, Rv1504c and Rv3845-3846 that are not present in L2 (74).

Another point of discussion is whether the insertion sites of IS*6110* occurs at insertional hotspots, i.e. independent insertion at the exact same point in different strains, with little or no phylogenetic signal, or inherited and relatively stable over time with significant phylogenetic signal. In this study, it was found that some Beijing strains have high copy numbers of IS*6110* whereas the others have low copy numbers as shown in appendix 2. It might be possible that the rate of transposition is high in some of the strains compared to the others as reported earlier by Dale et al. (107). This was not studied in this thesis but could be investigated in future works. It also has been shown by Dale et al., that these insertions of IS*6110* do not occur at hot spots but the transposition of IS*6110* is slow in low copy number strains than high copy strains (107). Roychowdhury et al supported the previous work of this thesis in that the IS*6110* hotspots observed are unique insertions subsequently inherited by daughter strains showing vertical transmission within genetic lineages, and confirmed by WGS analysis.

Following mapping of insertion sites, the IS*6110* FAFLP assay paved the way to initiate the development of a rapid PCR assay that could identify the TB genetic lineages directly from DNA via identification of lineage specific IS*6110* insertion points, without the need for performing FAFLP. This tool could rapidly assign lineages to the strain collection in any geographical setting with a basic PCR set up. The IS*6110* FAFLP method, though relatively simple and robust, would still be challenging to implement in settings with very poor resources, as the procedure has multiple steps, needs good quality DNA to be extracted from isolates, uses relatively expensive molecular biology reagents and equipment and also requires a cold chain infrastructure. The development of lineage specific PCR was important so that identification of genetic lineages directly from sputum samples would be possible (not tested in this project). This assay will not directly affect the treatment regimen of a patient but it will give an informed decision about the TB lineage prevalent and could help to direct the control measures for that particular lineage. It has been noticed that in certain regions like India, Pakistan, China, certain TB lineages (CAS-Delhi in India and Pakistan, Beijing in China) are more prevalent than other lineages (163,164) and does not necessarily correlate with the increased drug resistance patterns seen in these lineages (204) but could be to do with either a high density of human population in these areas (205) or the ineffective or mismanagement of the antibiotic usage in these countries (23).

In chapter 4, isolates representative of PGG3 and PGG1 lineages were selected and tested. This PCR method with specific primers targeting unique IS*6110* fragments was able to identify PGG3 and PGG1 genetic lineages, by detecting the unique 296bp product of PGG3 (360 R fragment inserted into oxidoreductase as shown in chapter 4, table 4-2) and a 200bp product of PGG1. Thus, the proof of principle was established and it would be useful to design PCRs to common fragments in the remaining lineages.

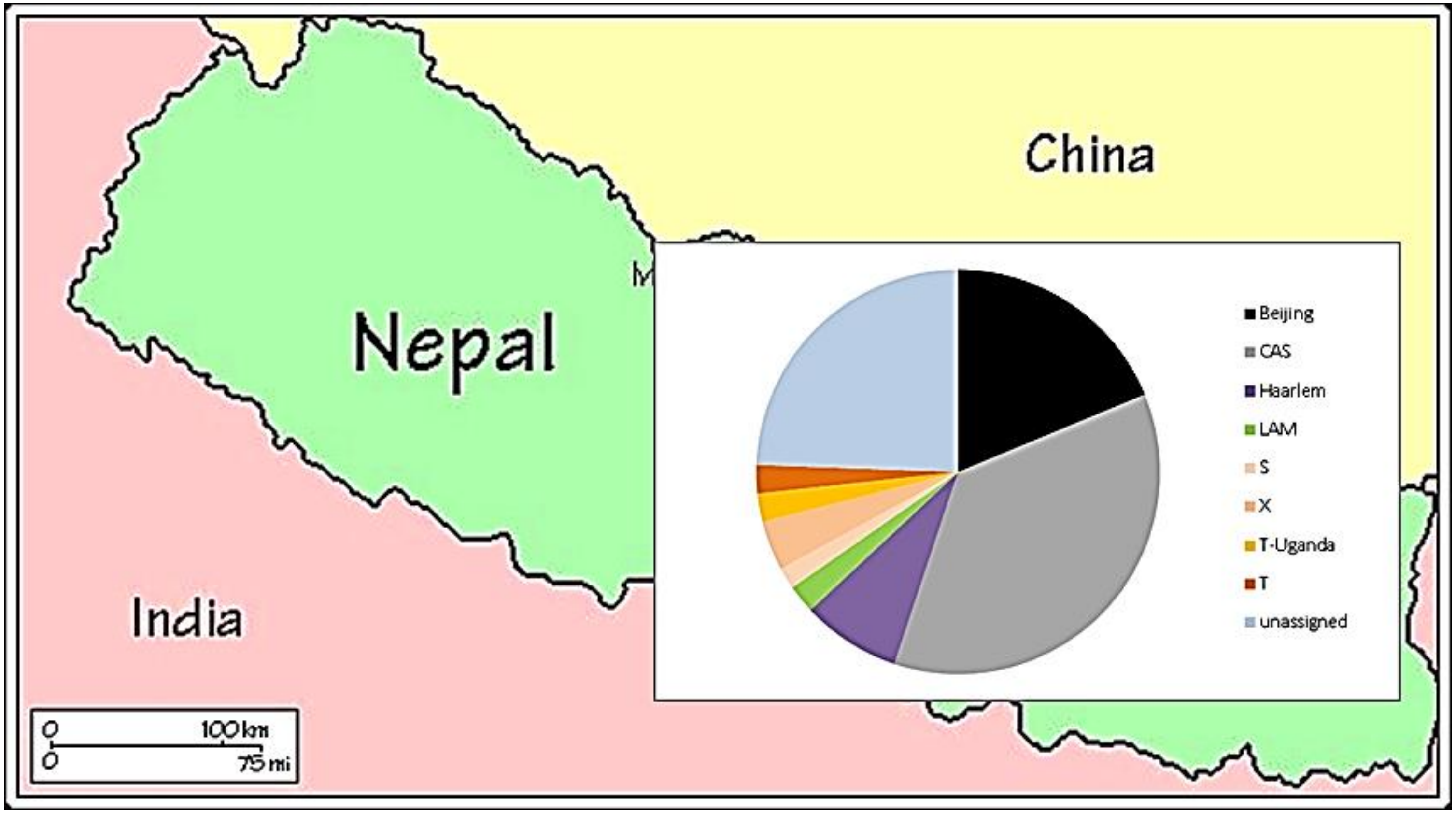
7.2.3 Classification of Nepalese TB clinical isolates into different TB genetic lineages (published work - (188))

The main function of this genomic tool is to characterise the TB strains studied into different TB genetic lineages from any geographical location. This technique was applied on readily available TB DNA samples from Nepal as shown in chapter 5 as there was limited or no information present on the genetic diversity of the circulating TB lineages in Nepal when this study started in 2011. However, a study published by Malla et al., in 2012, described the initial characterisation of these lineages and the importance of understanding their distribution in Nepal (191). Nepal's position geographically, sharing its borders between India and China, also proved relevant as one third of the total human TB cases are from these two countries (23). This emphasised that Nepal's location would have acted as a mixing pot for the strains originating from infections in India and China, reflected in our results in chapter 6 that 55% of the strains were either Beijing or CAS belonging to PGG1 group (188).

The findings of this study corroborates Malla's study (191) in that the majority of strains circulating in this geographical location in Nepal are from Lineages 2 and 3, PGG1- CAS and Beijing. These strains are predominantly found in India (CAS) and China (Beijing). Migration of the human population carrying these strains from these countries would have settled in Nepal and mixed with the local population thereby passing on these successful TB lineages as discussed in chapter 5). In chapter 5, IS*6110* FAFLP derived phylogeny also clearly demonstrates that in the majority of Nepalese TB patients, although usually infected with strains belonging to PGG1, can also be infected with a range of genetic lineages. This scenario clearly shows that various TB genetic lineages have entered Nepal at some point of time and the most successful lineages among them are the Beijing and the CAS lineages. This again focuses the discussion on human migration between the open borders of India in the south and China in the north of Nepal. In Northern India, the CAS- Delhi lineages are commonly seen and in China, Beijing lineages are predominant and this is evident in this study too (163,164,190). In chapter 5, the distribution of Beijing and CAS lineages are shown as 18.8 % and 36.6% respectively as seen in table 5-1. The perfect scenario is being offered by Nepal for the adaptation of *M. tuberculosis* strains by means of human migration for the purposes of tourism, religious travel like the pilgrimage to the famous religious shrines, work, and also cheaper treatment facilities (190).

In chapter 5, it was shown by IS*6110* FAFLP that there are a high number of strains belonging to the CAS lineages compared to the Beijing lineages which was also the case in Malla's study (191) where CAS lineages were more predominant than Beijing lineages. In Northern India, the predominant lineage seen is CAS and Beijing in China as shown in the figure 7-1 which shows that the geographical location of Nepal has played an important role in the TB lineages' distribution in Nepal (189), by being a perfect geographical location for the intermingling of different lineages from these two high burden countries.

A.


China
Nepal
India
0 100 km
0 75 mi
Beijing
CAS
Haarlem
LAM
S
X
T-Uganda
T
unassigned

B.

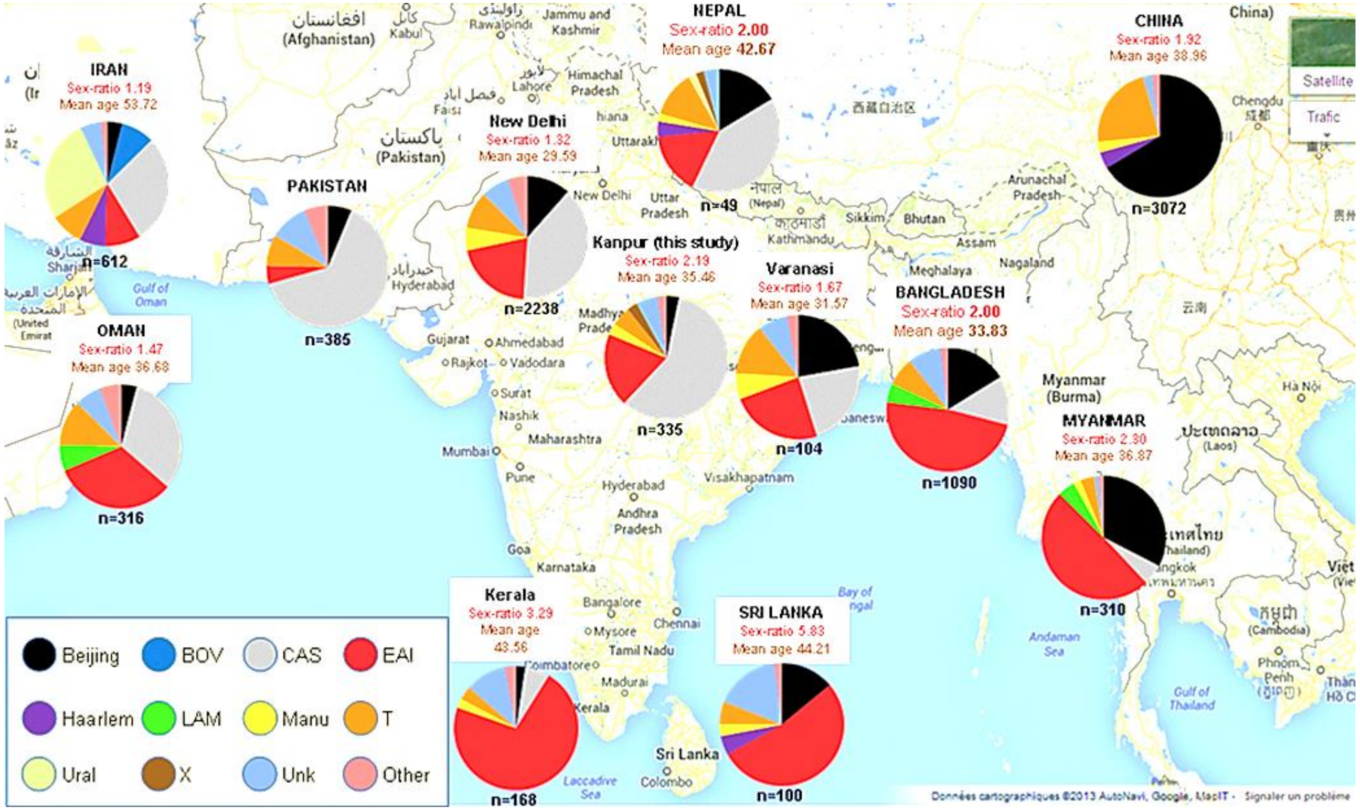


Figure 7-1 A. **Distribution of different TB lineages in Nepal from this study and** B. **the distribution of different lineages seen in India, China and other countries near the Indian subcontinent**

(this figure has been adapted from (189)).

7.2.4 Rifampicin Resistance status in Nepalese TB isolates from clinical samples (published work - (188,195).

Finally, as shown in chapter 6 by the *rpoB* PCR assay, seven of the 176 isolates were found to be likely resistant to rifampicin and all the seven mutations identified were in the RRDR region of *rpo*B. They were found to be non-synonymous thereby likely affecting the sensitivity of the strain. All 176 isolates in this study were primary untreated cases and this incidence of resistance seemed to be higher (3.9%) than the average national average of new, untreated cases (2.2%) (173,188). However the result was concordant with the recent study by Creswell et al., that showed the resistance as 3.3% (198) by analysing the resistance patterns using GeneXpert MTB/RIF kits. Again the increase in drug resistant TB in Nepal is probably due to the increase in drug resistant strains in India and China. The most frequently mutated codon is 531 followed by 526, 522 and 516 (see table 6-1), as observed by Poudel et al in his study (190). In Northern India, it has been reported that the majority of rifampicin mutation occur at codon 531 situated in the RRDR region (53,206) which reinforces the fact that the movement of human population helps the transmission of TB (159). One drawback of this technique is that it tests resistance to rifampicin by analysis of mutation in the RRDR only, even though sequencing of this region of *rpo*B has been suggested as a surrogate marker for MDR (53,196,207,208), it will not pick up 100% of mutations associated with resistance to this drug. This strategy will not be as useful for successful treatment of the patients if they are multi-resistant to other TB drugs and ideally the inclusion of a test to identify *ka*tG 315, the marker most commonly associated with isoniazid resistance, and other front line drugs, could also be included in a PCR screen for optimum assay sensitivity directly from samples.

Nepal, the main setting of this study is no longer classified as a high burden TB country (172). However it is considered as a resource limited setting due to the poor infrastructure in remote villages and the lack of technologically advanced diagnostic assays and skilled labour for analysing the results, thus preventing routine molecular analysis. The priority of these settings is to identify tuberculosis infection by collecting samples for bacterial isolation and to inform treatment. The period of time taken to identify TB and start patient specific therapy is where these settings fail to keep up pace with a resourced setting. An assay such as IS*6110* FAFLP PCR and, in particular the rapid PCR lineage detection method, although not a TB diagnostic test, could be useful to characterise TB isolates into different TB lineages directly from the clinical sample and thereby aid in contact tracing and indirectly in the control of outbreak. The resources needed are thermal cyclers, PCR reagents and trained staff for analysis and interpretation of results, which are already available in these settings. As contamination of cultures is potentially a major issue in outbreak investigation, the highly specific and sensitive nature of the assay could aid epidemiological investigations if implemented. The rapid PCR assay described in chapter 5 identifies lineage related strains and, with an expanded panel of PCRs for epidemiological investigations, control measures can be implemented within 2-3 days. If more information is needed on drug sensitivity then additional tests could also be added to this assay.

Novel diagnostic methods and assays are always needed for the effective and timely treatment of tuberculosis in resource limited settings. As the expenses incurred by the patients and their families accounted to 53% of annual household income per capita (209), the treatment of TB patients becomes extremely challenging in low income/high burden TB countries.

Different technologies evolved during the timeframe of this thesis work. WGS can now be performed relatively easily in high resource settings. Within a short period of time, strains could be identified and characterised using NGS technologies. In terms of rapid diagnosis however no other technology apart from single molecule WGS has shown promise in this area. Oxford Nanopore’s MinION sequencing could sequence TB genomic DNA directly from samples following extraction and depletion of human DNA. Currently however there is still an issue with the cost and availability of this technique in resource limited settings (148).

The advantages and disadvantages of the IS*6110* FAFLP genomic tool are summarised in the table 7-1.

Table 7-1 Brief summary of the advantages and disadvantages of IS*6110* FAFLP

Advantages	**Disadvantages**
High resolution marker and precise as 0.5bp difference can be a different lineage or strain (116.R is CAS and 117.1 R is Haarlem- table 5-1 and figure 5-1)	Difficult to characterise strains with less than 5 IS*6110* copies as shown in table 5-1 and figure 5-1 in chapter 5 where nearly 24% of the Nepalese samples were ungrouped and warrants another high resolution marker like MIRU-VNTR.
Simple, reproducible and robust technique than RFLP.	Although more rapid than RFLP, it can take up to several days to perform the technique, get the results and analyse them but the rapid lineage specific PCR described in chapter 4 overcomes this and the requirement for multiple enzymatic steps on extracted DNA from cultures.
Variations of the assay can suit both a high throughput modern lab and a resource poor setting depending on the equipment available	Basic equipment like PCR machines, heat blocks, gel electrophoresis and centrifuges are still needed, even if rapid tests are employed.
Continuous screening of the samples using the technique would aid contact tracing for immediate epidemiological intervention	Additional sequencing or other mutational studies like the rpoB PCR assay used in chapter 6 needed as this study does not detect any mutational changes like SNP.

7.3 Future work

A full toolbox of lineage specific PCRs could be developed and multiplex panels created. WGS may offer more information regarding the biological impact of IS*6110* insertion. Some insertion sites have remained in lineages for thousands of years and it is likely that they may confer a biological advantage. Knowing that they are present in specific locations could inform disease progression or pathogenicity in certain demographic groups. Characterisation of these and development of rapid PCRs will be useful for intervention. In terms of Nepal's story, it would be interesting to study the pattern of resistance in Nepal's TB strains and compare the resistance between the bordering countries. This will help in understanding the measures taken by the organism in developing resistance in a new niche. Also, it would be important to study the effect of different IS*6110* copy numbers in different TB lineages and their pathogenicity and if there is a relation between the two concepts.

7.4 Summary

The findings of all chapters have been summarised and tabulated below in table 7-2.

Table 7-2 Summary of the results from all chapters in this thesis

Chapter Headings	Aims	Approach	Main Findings	Page Numbers
1. Introduction	To set the scene for the thesis work	Extensive review of the literature	-Several epidemiological and genetic markers available to type TB. -Among others, IS*6110* found to contain phylogenetic signal and could therefore be used to develop a genomic tool box for resource poor settings.	1-62
2. General Materials and Methods	To document the standard methods used throughout the thesis.	Systematic approach from conception of idea, followed by development and implementation of the methods to achieve the	IS*6110* FAFLP PCR method using specific primers found to be the principal tool used in the project.	63-72

		primary aim of the project.		
3. Optimisation and Development of IS*6110* FAFLP PCR	To develop and standardise IS*6110* FAFLP PCR that is both sensitive and specific.	Test the samples using different conditions and reagents to achieve the desired result.	The method was standardised and was ready to be validated.	73-91
4. Mapping of Insertion sites (IS*6110*) in the M. tuberculosis H37Rv reference genome and rapid definition of genetic lineage	- To prove that IS*6110* FAFLP PCR is accurate and could map insertion sites in the completely sequenced H37Rv strain. - To achieve the aim of developing a rapid lineage specific PCR	- H37Rv was subjected to IS*6110* FAFLP and then insertion sites were mapped using two base selective primers. - Primers designed for the PGG1 and PGG3 strains were used in PCR reactions and tested for specificity	- Sixteen insertion sites were mapped onto the H37Rv genome, thereby proving the accuracy of the technique. - Proof of principle was established by detecting 100bp and 296bp products specific for PGG1 and PGG3 groups respectively	92-107 (published) (179) .
5. Classification of Nepalese TB clinical	To characterise the TB strains from Nepal, a	Bacterial DNA from all 176 TB isolates was subjected to	-All 176 isolates were delineated into different TB	108-118 (published)

isolates into different TB genetic lineages	resource poor setting, using IS*6110* FAFLP PCR.	IS*6110* FAFLP PCR and then characterised using BioNumerics software.	genetic lineages. -55% of the strains belonged to PGG1 group, either Beijing or CAS strains.	(188)
6. Rifampicin Resistance status in Nepalese TB isolates from clinical samples	To infer the drug-resistance status in the Nepalese TB samples.	Using *rpo*B resistant PCR, all 176 samples were checked for inferred drug-resistance by observing the mutations in the RRDR region of the *rpo*B gene.	-Seven isolates were found to have mutations at the normally reported mutation sites. -3.9% of the strains were found to be MDR-TB -Prevalence of MDR-TB is high - Results concordant with previously published results	119-125 (published) (188,195)

7. Final Discussion	To contextualise the results obtained and make judgement of the aims of the thesis.	Summarise the results obtained and discuss the relevance of the results with respect to the aim and objectives of the thesis.	-Summarised the results of all the chapters. -Discussion of different results sections in detail. -Showed the efficacy of the IS*6110* FAFLP PCR assay as a genomic tool in a resource poor setting.	126-146

7.5 Conclusions

The aim of this thesis was to develop a genomic mapping tool using IS*6110* FAFLP that could be used in resource poor settings as an assay for the characterisation of *M. tuberculosis* strains. This aim was achieved by first optimising the assay (chapter 3), testing it on H37Rv (chapter 4) by mapping the IS*6110* insertion sites followed by the delineation of Nepalese TB strains into different TB lineages as shown in chapter 5. The IS*6110* FAFLP PCR data led to the development of the rapid lineage specific PCR assay (chapter 4) that can identify the circulating TB genotypes in a specific geographic area. Though a rapid PCR assay was not developed for all the lineages or genotypes, the proof of principle was established by the identification of PGG1 and PGG3 strains. This work has successfully developed a simple, rapid and robust method to characterise different strains of *M. tuberculosis* complex that could aid and inform epidemiological investigation and intervention that could be used in resource poor settings.

References

1. WHO. Global Tuberculosis Report 2017. 2017. 28-58 p.

2. Magee J, Ward A. Genus I. *Mycobacterium*. Bergey's Manual of Systematic Bacteriology, The Actinobacteria M Goodfellow, P Kämpfer, HJ Busse, ME Trujillo, KI Suzuki, W Ludwig & WB Whitman (eds) New York,: Springer. 2012. p. 312–75.

3. Alexander KA, Laver PN, Michel AL, Williams M, van Helden PD, Warren RM, et al. Novel *Mycobacterium tuberculosis* complex pathogen, M. mungi. Emerg Infect Dis. 2010 Aug;16(8):1296–9.

4. Dippenaar A, Parsons SDC, Sampson SL, van der Merwe RG, Drewe JA, Abdallah AM, et al. Whole genome sequence analysis of *Mycobacterium suricattae*. Tuberculosis. 2015;1–7.

5. Huard RC, Fabre M, de Haas P, Lazzarini LCO, van Soolingen D, Cousins D, et al. Novel genetic polymorphisms that further delineate the phylogeny of the *Mycobacterium tuberculosis* complex. J Bacteriol. 2006;188(12):4271–87.

6. Boritsch EC, Supply P, Honoré N, Seeman T, Stinear TP, Brosch R. A glimpse into the past and predictions for the future: The molecular evolution of the tuberculosis agent. Mol Microbiol. 2014 Jul 1;

7. Laënnec RTH. De l'auscultation médiate ou Traité du Diagnostic des

Maladies des Poumon et du Coeur. 1st ed. Paris: Chez J.A. Brosson & J. S. Chaudé; 1819. 530 p.

8. Villemin JA. Études sur la tuberculose : preuves rationnelles et expérimentales de sa spécificité et de son inoculabilité. Paris: J.B. Bailliere et fils; 1868. 662 p.

9. Schluger NW. The pathogenesis of tuberculosis: the first one hundred (and twenty-three) years. Am J Respir Cell Mol Biol. 2005;32(4):251–6.

10. Koch R. Die Aetiologie der Tuberculose. Berliner Klin Wochenschrift. 1882;15.

11. Lienhardt C, Glaziou P, Uplekar M, Lönnroth K, Getahun H, Raviglione M. Global tuberculosis control: lessons learnt and future prospects. Nat Rev Microbiol. 2012;10(6):407.

12. WHO. Global Tuberculosis report 2016. 2016;

13. Knechel NA. Tuberculosis: pathophysiology, clinical features, and diagnosis. Crit Care Nurse. 2009;29(2):34–43.

14. Lee REB, Li W, Chatterjee D, Lee RE. Rapid structural characterization of the arabinogalactan and lipoarabinomannan in live mycobacterial cells using 2D and 3D HR-MAS NMR: structural changes in the arabinan due to ethambutol treatment and gene mutation are observed. Glycobiology. 2005;15(2):139–51.

15. Joe M, Bai Y, Nacario RC, Lowary TL. Synthesis of the docosanasaccharide arabinan domain of mycobacterial arabinogalactan and a proposed octadecasaccharide biosynthetic precursor. J Am Chem Soc. 2007;129(32):9885–901.

16. Dunlap N, Bass J, Fujiwara P, Hopewell P, Horsburgh Jr C, Salfinger M, et al. Diagnostic Standards and Classification of Tuberculosis in Adults and Children. Am J Respir Crit Care Med. 2000;161:1376–95.

17. Wurie FB, Lawn SD, Booth H, Sonnenberg P, Hayward AC. Bioaerosol production by patients with tuberculosis during normal tidal breathing: implications for transmission risk. Thorax. 2016;(71):549–54.

18. Lee SH. Tuberculosis Infection and Latent Tuberculosis. Tuberc Respir Dis (Seoul). 2016;79(4):201–6.

19. Li Y, Petrofsky M, Bermudez LE, Mmun INI. *Mycobacterium tuberculosis* Uptake by Recipient Host Macrophages Is Influenced by Environmental Conditions in the Granuloma of the Infectious Individual and Is Associated with Impaired Production of Interleukin-12 and Tumor Necrosis Factor Alpha. Infect Immun. 2002;70(11):6223–30.

20. Dheda K, Booth H, Huggett JF, Johnson M a, Zumla A, Rook G a W. Lung remodeling in pulmonary tuberculosis. J Infect Dis. 2005;192(7):1201–9.

21. O’Connor JA, O’Reilly B, Corcoran GD, O’Mahony J, Lucey B.

Mycobacterium diagnostics: from the primitive to the promising. Br J Biomed Sci. 2015;72(1):32–41.

22. Pai M, Schito M. Tuberculosis Diagnostics in 2015: Landscape, Priorities, Needs, and Prospects. J Infect Dis. 2015;211(suppl 2):S21–8.

23. WHO. Global tuberculosis report 2014 (WHO/HTM/TB/2014.08). 2014;

24. Bomanji JB, Gupta N, Gulati P, Das CJ. Imaging in Tuberculosis. Cold Spring Harb Perspect Med. 2015;5(6):a017814–a017814.

25. Skoura E, Zumla A, Bomanji J. Imaging in tuberculosis. Int J Infect Dis. 2015;32:87–93.

26. Lalvani A, Pareek M. Interferon gamma release assays: principles and practice. Enferm Infecc Microbiol Clin. 2010;28(4):245–52.

27. Dye C, Scheele S, Dolin P, Pathania V, Raviglione MC. Consensus statement. Global burden of tuberculosis: estimated incidence, prevalence, and mortality by country. WHO Global Surveillance and Monitoring Project. JAMA. 1999;282(7):677–86.

28. Behr MA, Warren SA, Salamon H, Hopewell PC, Ponce de Leon A, Daley CL, et al. Transmission of *Mycobacterium tuberculosis* from patients smear-negative for acid-fast bacilli. Lancet (London, England). 1999;353(9151):444–9.

29. Steingart KR, Ng V, Henry M, Hopewell PC, Ramsay A, Cunningham J, et al. Sputum processing methods to improve the sensitivity of smear microscopy for tuberculosis: a systematic review. Lancet Infect Dis. 2006;6(10):664–74.

30. WHO. Fluorescent light-emitting diode (LED) microscopy for diagnosis of tuberculosis Policy statement. WHO. 2011;(March):1–12.

31. Corbett EL, Watt CJ, Walker N, Maher D, Williams BG, Raviglione MC, et al. The growing burden of tuberculosis: global trends and interactions with the HIV epidemic. Arch Intern Med. 2003;163(9):1009.

32. Hargreaves N, Kadzakumanja O. “Smear-negative”pulmonary tuberculosis in a DOTS programme: poor outcomes in an area of high HIV seroprevalence. Int J Tuberc Lung Dis. 2001;5(May):847–54.

33. Chihota VN, Grant AD, Fielding K, Ndibongo B, van Zyl A, Muirhead D, et al. Liquid vs. solid culture for tuberculosis: performance and cost in a resource-constrained setting. Int J Tuberc lung Dis. 2010;14(8):1024–31.

34. WHO. Global Tuberculosis Report. 2012;

35. Langley I, Lin H-H, Egwaga S, Doulla B, Ku C-C, Murray M, et al. Assessment of the patient, health system, and population effects of Xpert MTB/RIF and alternative diagnostics for tuberculosis in Tanzania: an integrated modelling approach. Lancet Glob Heal. 2014;2(10):e581–

91.

36. Viveiros M, Leandro C, Rodrigues L. Probe assay for rapid identification of *Mycobacterium tuberculosis* complex strains and detection of rifampin resistance in 360 smear-positive respiratory specimens. J Clin Microbiol. 2005;43(9):4880–4.

37. Drobniewski F, Nikolayevskyy V, Maxeiner H, Balabanova Y, Casali N, Kontsevaya I, et al. Rapid diagnostics of tuberculosis and drug resistance in the industrialized world: clinical and public health benefits and barriers to implementation. BMC Med. 2013;11(1):190.

38. Mironova S, Pimkina E, Kontsevaya I, Nikolayevskyy V, Balabanova Y, Skenders G, et al. Performance of the GenoType® MTBDRPlus assay in routine settings: a multicenter study. Eur J Clin Microbiol Infect Dis. 2012;31(7):1381–7.

39. Nikolayevskyy V, Balabanova Y, Simak T, Malomanova N, Fedorin I, Drobniewski F. Performance of the Genotype MTBDRPlus assay in the diagnosis of tuberculosis and drug resistance in Samara, Russian Federation. BMC Clin Pathol. 2009;9:2.

40. Kontsevaya I, Ignatyeva O, Nikolayevskyy V, Balabanova Y, Kovalyov A, Kritsky A, et al. Diagnostic accuracy of the genotype MTBDRsl assay for rapid diagnosis of extensively drug-resistant tuberculosis in HIV-coinfected patients. J Clin Microbiol. 2013;51(1):243–8.

41. Brown AC, Bryant JM, Einer-Jensen K, Holdstock J, Houniet DT, Chan JZM, et al. Rapid Whole Genome Sequencing of *M. tuberculosis* directly from clinical samples. J Clin Microbiol. 2015;44(May):2230–7.

42. Köser CU, Bryant JM, Becq J, Török ME, Ellington MJ, Marti-Renom MA, et al. Whole-Genome Sequencing for Rapid Susceptibility Testing of *M. tuberculosis*. N Engl J Med. 2013;369(3):290–2.

43. Lewinsohn DM, Leonard MK, LoBue PA, Cohn DL, Daley CL, Desmond E, et al. Official American Thoracic Society/Infectious Diseases Society of America/Centers for Disease Control and Prevention Clinical Practice Guidelines: Diagnosis of Tuberculosis in Adults and Children. Clin Infect Dis. 2017;64(2):111–5.

44. Ahmad S, Mokaddas E. Current status and future trends in the diagnosis and treatment of drug-susceptible and multidrug-resistant tuberculosis. J Infect Public Health. 2014;7(2):75–91.

45. Pai M, Kalantri S, Dheda K. New tools and emerging technologies for the diagnosis of tuberculosis: Part II. Expert Rev Mol Diagn. 2006;6(3):423–32.

46. Dheda K, Barry CE, Maartens G. Tuberculosis. Lancet. 2016;387(10024):1211–26.

47. WHO. Global Tuberculosis Control 2011. 2011.

48. Fonseca JD, Knight GM, McHugh TD. The complex evolution of

antibiotic resistance in *Mycobacterium tuberculosis*. Int J Infect Dis. 2015;32:94–100.

49. Blumberg HM, Burman WJ, Chaisson RE, Daley CL, Etkind SC, Friedman LN, et al. American Thoracic Society/Centers for Disease Control and Prevention/Infectious Diseases Society of America: treatment of tuberculosis. Am J Respir Crit Care Med. 2003;167(4):603–62.

50. Mitchison D, Davies G. The chemotherapy of tuberculosis: past, present and future. Int J Tuberc Lung Dis. 2012;16(6):724–32.

51. Almeida Da Silva PE, Palomino JC. Molecular basis and mechanisms of drug resistance in *Mycobacterium tuberculosis*: classical and new drugs. J Antimicrob Chemother. 2011;66(7):1417–30.

52. Telenti A, Imboden P, Marchesi F, Lowrie D, Cole S, Colston MJ, et al. Detection of rifampicin-resistance mutations in *Mycobacterium tuberculosis*. Lancet. 1993;341(8846):647–50.

53. Suresh N, Singh UB, Arora J, Pande JN, Seth P, Samantaray JC. Rapid detection of rifampicin-resistant *Mycobacterium tuberculosis* by in-house, reverse line blot assay. Diagn Microbiol Infect Dis. 2006;56(2):133–40.

54. Al-Mutairi NM, Ahmad S, Mokaddas E. Performance comparison of four methods for detecting multidrug-resistant *Mycobacterium*

tuberculosis strains. Int J Tuberc lung Dis. 2011;15(1):110–5.

55. Billington OJ, McHugh TD, Gillespie SH. Physiological cost of rifampin resistance induced in vitro in *Mycobacterium tuberculosis*. Antimicrob Agents Chemother. 1999;43(8):1866–9.

56. Zhang Y, Heym B, Allen B, Young D, Cole S. The catalase—peroxidase gene and isoniazid resistance of *Mycobacterium tuberculosis*. Nature. 1992;358(6387):591–3.

57. Ahmad S, Mokaddas E. Contribution of AGC to ACC and other mutations at codon 315 of the katG gene in isoniazid-resistant *Mycobacterium tuberculosis* isolates from the Middle East. Int J Antimicrob Agents. 2004;23(5):473–9.

58. O'Sullivan DM, McHugh TD, Gillespie SH. The effect of oxidative stress on the mutation rate of *Mycobacterium tuberculosis* with impaired catalase/peroxidase function. J Antimicrob Chemother. 2008;62(4):709–12.

59. Scorpio A, Zhang Y. Mutations in *pncA*, a gene encoding pyrazinamidase/nicotinamidase, cause resistance to the antituberculous drug pyrazinamide in tubercle bacillus. Nat Med. 1996;2(6):662–7.

60. Safi H, Lingaraju S, Amin A, Kim S, Jones M, Holmes M, et al. Evolution of high-level ethambutol-resistant tuberculosis through

interacting mutations in decaprenylphosphoryl-β-D-arabinose biosynthetic and utilization pathway genes. Nat Genet. 2013;45(10):1190–7.

61. Gillespie SH, Crook AM, McHugh TD, Mendel CM, Meredith SK, Murray SR, et al. Four-Month Moxifloxacin-Based Regimens for Drug-Sensitive Tuberculosis. N Engl J Med. 2014;371(17):1577–87.

62. Rieder HL. Fourth-generation fluoroquinolones in tuberculosis. Lancet. 2009;373(9670):1148–9.

63. Ravenel MP. La Vaccination Préventive Contre la Tuberculose par le “BCG.” Am J Public Heal Nations Heal. 1928;18(8):1075.

64. Andersen P, Kaufmann SHE. Novel Vaccination Strategies against Tuberculosis. Cold Spring Harb Perspect Med. 2014;4(6):a018523.

65. Moliva JI, Turner J, Torrelles JB. Prospects in *Mycobacterium bovis* Bacille Calmette et Guérin (BCG) vaccine diversity and delivery: Why does BCG fail to protect against tuberculosis? Vaccine. 2015;33(39):5035–41.

66. Moliva JI, Turner J, Torrelles JB. Immune Responses to Bacillus Calmette-Guérin Vaccination: Why Do They Fail to Protect against *Mycobacterium tuberculosis*? Front Immunol. 2017;8:407.

67. Rodrigues LC, Diwan VK, Wheeler JG. Protective effect of BCG against tuberculous meningitis and miliary tuberculosis: a meta-

analysis. Int J Epidemiol. 1993;22(6):1154–8.

68. Tameris MD, Hatherill M, Landry BS, Scriba TJ, Snowden MA, Lockhart S, et al. Safety and efficacy of MVA85A, a new tuberculosis vaccine, in infants previously vaccinated with BCG: a randomised, placebo-controlled phase 2b trial. Lancet. 2013;381(9871):1021–8.

69. Thorne N, Underwood A, Gharbia S, Arnold C. Evolutionary clues from comparative analysis of *Mycobacterium tuberculosis* variable-number tandem repeat sequences within genetic families. Infect Genet Evol. 2007a;7(2):239–46.

70. Hermans PWM, van Soolingen D, Bik EM, DeHaas PEW, Dale JW, VanEmbden J. Insertion Element IS*987* from *Mycobacterium bovis* BCG Is Located in a Hot-Spot Integration Region for Insertion Elements in *Mycobacterium tuberculosis* Complex Strains. Infect Immun. 1991;59(8):2695–705.

71. Hermans PW, van Soolingen D, Dale JW, Schuitema AR, McAdam RA, Catty D, et al. Insertion element IS*986* from *Mycobacterium tuberculosis*: a useful tool for diagnosis and epidemiology of tuberculosis. J Clin Microbiol. 1990;28(9):2051–8.

72. McAdam RA, Hermans, P .W M, van Soolingen D, Zainuddin ZF, Catty D, van Embden, J DA, et al. Characterization of a *Mycobacterium tuberculosis* insertion sequence belonging to the IS*3* family. Mol Microbiol. 1990;4(9):1607–13.

73. Simonet M, Riot B, Fortineau N, Berche P. Invasin production by *Yersinia pestis* is abolished by insertion of an IS*200*-like element within the inv gene. Infect Immun. 1996;64(1):375–9.

74. Roychowdhury T, Mandal S, Bhattacharya A. Analysis of IS*6110* insertion sites provide a glimpse into genome evolution of *Mycobacterium tuberculosis*. Sci Rep. 2015;5(July):12567.

75. Delihas N. Impact of small repeat sequences on bacterial genomes and genome evolution. Genome Biol Evol. 2011;3:959–73.

76. Park I, Saurin W, Ullmann A. A highly conserved 530 base-pair repeated DNA sequence specific for *Bordetella pertussis*. FEMS Microbiol Lett. 1988;52:19–24.

77. Hohn B, Collins J. A small cosmid for efficient cloning of large DNA fragments. Gene. 1980;11:291–8.

78. Thierry D, Cave MD, Eisenach KD, Crawford JT, Bates JH, Gicquel B, et al. IS*6110*, an IS-like element of *Mycobacterium tuberculosis* complex. Nucleic Acids Res. 1990;18(1):188.

79. Van Duin JM, Pijnenburg JEM, van Rijswoud CM, de Haas PEW, Hendriks WDH, van Soolingen D. Investigation of cross contamination in a *Mycobacterium tuberculosis* laboratory using IS*6110* DNA fingerprinting. International journal of tuberculosis and lung disease. 1998. p. 425–9.

80. Cave MD, Eisenach KD, McDermott PF, Bates JH, Crawford JT. IS*6110*: conservation of sequence in the *Mycobacterium tuberculosis* complex and its utilization in DNA fingerprinting. Mol Cell Probes. 1991;5(1):73–80.

81. Gordon S V, Heym B, Parkhill J, Barrell B, Cole ST. New insertion sequences and a novel repeated sequence in the genome of *Mycobacterium tuberculosis* H37Rv. Microbiology. 1999;145:881–92.

82. Mendiola M V, Martín C, Otal I, Gicquel B. Analysis of the regions responsible for IS*6110* RFLP in a single *Mycobacterium tuberculosis* strain. Res Microbiol. 1992;143(8):767–72.

83. Iovannisci DM, Winn-Deen ES. Ligation amplification and fluorescence detection of *Mycobacterium tuberculosis* DNA. Mol Cell Probes. 1993;7(1):35–43.

84. Haas WH, Butler WR, Woodley CL, Crawford JT. Mixed-Linker Polymerase Chain Reaction : a New Method for Rapid Fingerprinting of Isolates of the *Mycobacterium tuberculosis* Complex. Tuberculosis. 1993;31(5):1293–8.

85. Plikaytis BB, Crawford JT, Woodley CL, Butler WR, Eisenach KD, Cave MD, et al. Rapid, amplification-based fingerprinting of *Mycobacterium tuberculosis*. J Gen Microbiol. 1993;139(7):1537–42.

86. Niemann S, Köser CU, Gagneux S, Plinke C, Homolka S, Bignell H, et

al. Genomic diversity among drug sensitive and multidrug resistant isolates of *Mycobacterium tuberculosis* with identical DNA fingerprints. PLoS One. 2009;4(10):e7407.

87. Collins DM, Erasmuson SK, Stephens DM, Yates GF, De Lisle GW. DNA fingerprinting of *Mycobacterium bovis* strains by restriction fragment analysis and hybridization with insertion elements IS*1081* and IS*6110*. J Clin Microbiol. 1993;31(5):1143–7.

88. Zanini M, Moreira E, Lopes M, Oliveira R, Leão S, Fioravanti R, et al. *Mycobacterium bovis*: polymerase chain reaction identification in bovine lymphonode biopsies and genotyping in isolates from Southeast Brazil by spolygotyping and restriction fragment length polymorphism. Mem Inst Oswaldo Cruz. 2001;96(6):809–13.

89. Warren RM, Streicher EM, Sampson SL, Spuy GD Van Der, Richardson M, Nguyen D, et al. Microevolution of the Direct Repeat Region of *Mycobacterium tuberculosis* : Implications for Interpretation of Spoligotyping Data. J Clin Microbiol. 2002a;40(12):4457–65.

90. Warren RM, van der Spuy GD, Richardson M, Beyers N, Booysen C, Behr MA, et al. Evolution of the IS*6110*-based restriction fragment length polymorphism pattern during the transmission of *Mycobacterium tuberculosis*. J Clin Microbiol. 2002b;40(4):1277–82.

91. Warren RM, van der Spuy GD, Richardson M, Beyers N, Borgdorff MW, Behr MA, et al. Calculation of the stability of the IS*6110* banding

pattern in patients with persistent *Mycobacterium tuberculosis* disease. J Clin Microbiol. 2002c;40(5):1705–8.

92. Otal I, Martín C, Vincent-Lévy-Frebault V, Thierry D, Gicquel B. Restriction fragment length polymorphism analysis using IS*6110* as an epidemiological marker in tuberculosis. J Clin Microbiol. 1991;29(6):1252–4.

93. de Boer AS, Borgdorff MW, de Haas PE, Nagelkerke NJ, van Embden JD, van Soolingen D. Analysis of rate of change of IS*6110* RFLP patterns of *Mycobacterium tuberculosis* based on serial patient isolates. J Infect Dis. 1999;180(4):1238–44.

94. van Embden JDA, Cave MD, Crawford JT, Dale JW, Eisenach KD, Gicquel B, et al. Strain Identification of *Mycobacterium tuberculosis* by DNA Fingerprinting : Recommendations for a Standardized Methodology. J Clin Microbiol. 1993;31(2):406–9.

95. Kato-Maeda M, Metcalfe JZ, Flores L. Genotyping of *Mycobacterium tuberculosis*: application in epidemiologic studies. Future Microbiol. 2011;6(2):203–16.

96. Narayanan S. Molecular epidemiology of tuberculosis. Indian J Med Res. 2004;120(4):233–47.

97. Safi H, Barnes PF, Lakey DL, Shams H, Samten B, Vankayalapati R, et al. IS*6110* functions as a mobile, monocyte-activated promoter in

Mycobacterium tuberculosis. Mol Microbiol. 2004;52(4):999–1012.

98. Beggs ML, Eisenach KD, Cave MD. Mapping of IS*6110* insertion sites in two epidemic strains of *Mycobacterium tuberculosis*. J Clin Microbiol. 2000;38(8):2923–8.

99. Raimunda D, Long JE, Sassetti CM, Argüello JM. Role in metal homeostasis of *CtpD*, a Co^{2+} transporting P(1B4)-ATPase of *Mycobacterium smegmatis*. Mol Microbiol. 2012;84(6):1139–49.

100. Wall S, Ghanekar K, McFadden J, Dale JW. Context-sensitive transposition of IS*6110* in mycobacteria. Microbiology. 1999;145(Pt 1)(2):3169–76.

101. Fang Z, Forbes KJ. A *Mycobacterium tuberculosis* IS*6110* preferential locus (*ipl*) for insertion into the genome. J Clin Microbiol. 1997;35(2):479–81.

102. Fang Z, Doig C, Morrison N, Watt B, Forbes KJ. Characterization of IS*1547*, a new member of the IS*900* family in the *Mycobacterium tuberculosis* complex, and its association with IS*6110*. J Bacteriol. 1999;181:1021–4.

103. Kurepina NE, Sreevatsan S, Plikaytis BB, Bifani PJ, Connell ND, Donnelly RJ, et al. Characterization of the phylogenetic distribution and chromosomal insertion sites of five IS*6110* elements in *Mycobacterium tuberculosis*: non-random integration in the *dnaA-dnaN* region. Tuber

lung Dis. 1998;79(1):31–42.

104. Vera-Cabrera L, Hernández-Vera MA, Welsh O, Johnson WM, Castro-Garza J. Phospholipase region of *Mycobacterium tuberculosis* is a preferential locus for IS*6110* transposition. J Clin Microbiol. 2001;39(10):3499–504.

105. Sampson SL, Warren RM, Richardson M, van der Spuy GD, van Helden PD. Disruption of coding regions by IS*6110* insertion in *Mycobacterium tuberculosis*. Tuber lung Dis. 1999;79(6):349–59.

106. Yesilkaya H, Forbes KJ, Shafi J, Smith R, Dale JW, Rajakumar K, et al. The genetic portrait of an outbreak strain. Tuberculosis. 2006;86(5):357–62.

107. Dale JW, Al-Ghusein H, Al-Hashmi S, Butcher P, Dickens AL, Drobniewski F, et al. Evolutionary relationships among strains of *Mycobacterium tuberculosis* with few copies of IS*6110*. J Bacteriol. 2003;185(8):2555–62.

108. Mcadam RA, Quan S, Smith DA, Bardarov S, Betts JC, Cook FC, et al. Characterization of a *Mycobacterium tuberculosis* H37Rv transposon library reveals insertions in 351 ORFs and mutants with altered virulence. Microbiology. 2002;148(Pt 10):2975–86.

109. van Soolingen D, de Haas PE, Hermans PW, Groenen PM, van Embden JD. Comparison of various repetitive DNA elements as

genetic markers for strain differentiation and epidemiology of *Mycobacterium tuberculosis*. J Clin Microbiol. 1993;31(8):1987–95.

110. Ross BC, Raios K, Jackson K, Dwyer B. Molecular cloning of a highly repeated DNA element from *Mycobacterium tuberculosis* and its use as an epidemiological tool. J Clin Microbiol. 1992;30(4):942–6.

111. Chaves F, Yang Z, el Hajj H, Alonso M, Burman WJ, Eisenach KD, et al. Usefulness of the secondary probe pTBN12 in DNA fingerprinting of *Mycobacterium tuberculosis*. J Clin Microbiol. 1996;34(5):1118–23.

112. Kamerbeek J, Schouls LEO, Kolk A, Agterveld MVAN, Soolingen DVAN, Kuijper S, et al. Simultaneous Detection and Strain Differentiation of *Mycobacterium tuberculosis* for Diagnosis and Epidemiology. J Clin Microbiol. 1997;35(4):907–14.

113. 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;17:1–17.

114. Brosch R, Gordon S V, 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).

115. Ei PW, Aung WW, Lee JS, Choi GE, Chang CL. Molecular Strain Typing of *Mycobacterium tuberculosis*: a Review of Frequently Used

Methods. J Korean Med Sci. 2016;31(11):1673–83.

116. Filliol I, Motiwala AS, Cavatore M, Qi W, Hazbo MH, Bobadilla M, et al. Global Phylogeny of *Mycobacterium tuberculosis* Based on Single Nucleotide Polymorphism (SNP) Analysis : Insights into Tuberculosis Evolution , Phylogenetic Accuracy of Other DNA Fingerprinting Systems , and Recommendations for a Minimal Standard SNP Set. J Bacteriol. 2006;188(2):759–72.

117. Goyal M, Saunders NA, van Embden JD, Young DB, Shaw RJ. Differentiation of *Mycobacterium tuberculosis* isolates by spoligotyping and IS*6110* restriction fragment length polymorphism. J Clin Microbiol. 1997;35(3):647–51.

118. Streicher EM, Victor TC, Spuy G Van Der, Sola C, Rastogi N, Helden PD Van, et al. Spoligotype Signatures in the *Mycobacterium tuberculosis* Complex. J Clin Microbiol. 2007;45(1):237–40.

119. Demay C, Liens B, Burguière T, Hill V, Couvin D, Millet J, et al. SITVITWEB--a publicly available international multimarker database for studying *Mycobacterium tuberculosis* genetic diversity and molecular epidemiology. Infect Genet Evol. 2012;12(4):755–66.

120. Supply P, Magdalena J, Himpens S, Locht C. Identification of novel intergenic repetitive units in a mycobacterial two-component system operon. Mol Microbiol. 1997;26(5):991–1003.

121. Supply P, Mazars E, Lesjean S, Vincent V, Gicquel B, Locht C. Variable human minisatellite-like regions in the *Mycobacterium tuberculosis* genome. Mol Microbiol. 2000;36(3):762–71.

122. Frothingham R, Meeker-O’Connell WA. Genetic diversity in the *Mycobacterium tuberculosis* complex based on variable numbers of tandem DNA repeats. Microbiology. 1998;144 (Pt 5):1189–96.

123. Supply P, Allix C, Lesjean S, Cardoso-Oelemann M, Rüsch-Gerdes S, Willery E, 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.

124. Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, et al. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. 1995;23(21):4407–14.

125. Savelkoul P, Aarts H, de Haas J, Dijkshoorn B. Amplified-fragment length polymorphism analysis: the state of an art. J Clin Microbiol. 1999;37(10):3083–91.

126. Desai M, Tanna A, Wall R, Efstratiou A, George R, Stanley J. Fluorescent amplified-fragment length polymorphism analysis of an outbreak of group A streptococcal invasive disease. J Clin Microbiol. 1998;36(11):3133–7.

127. Arnold C, Metherell L, Willshaw G, Maggs A, Stanley J. Predictive

fluorescent amplified-fragment length polymorphism analysis of *Escherichia coli*: high-resolution typing method with phylogenetic significance. J Clin Microbiol. 1999;37(5):1274–9.

128. Meudt HM, Clarke AC. Almost Forgotten or Latest Practice ? AFLP applications, analyses and advances. Trends Plant Sci. 2007;12(3):106–17.

129. Bensch S, Akesson M. Ten years of AFLP in ecology and evolution: why so few animals? Mol Ecol. 2005;14(10):2899–914.

130. Goulding JN, Stanley J, Saunders N, Arnold C. Genome-sequence-based fluorescent amplified-fragment length polymorphism analysis of *Mycobacterium tuberculosis*. J Clin Microbiol. 2000;38(3):1121–6.

131. Randall MJ, Karst J, Pec GJ, Davis CS, Hall JC, Cahill JF. A molecular identification protocol for roots of boreal forest tree species. Appl Plant Sci. 2014;2(11):1400069.

132. Desai M, Logan JM, Frost JA, Stanley J. Genome sequence-based fluorescent amplified fragment length polymorphism of *Campylobacter jejuni*, its relationship to serotyping, and its implications for epidemiological analysis. J Clin Microbiol. 2001;39(11):3823–9.

133. Smith D, Willshaw G, Stanley J, Arnold C. Genotyping of verocytotoxin-producing *Escherichia coli* O157: comparison of isolates of a prevalent phage type by fluorescent amplified-fragment length polymorphism and

pulsed-field gel electrophoresis analyses. J Clin Microbiol. 2000;38(12):4616–20.

134. Amar C. Fluorescent amplified fragment length polymorphism (fAFLP) analysis of *Listeria monocytogenes*. Methods Mol Biol. 2014;1157:95–101.

135. Thorne N, Borrell S, Evans J, Magee J, García de Viedma D, Bishop C, et al. IS*6110*-based global phylogeny of *Mycobacterium tuberculosis*. Infect Genet Evol. 2011;11(1):132–8.

136. Palmer HM, Arnold C. Genotyping *Neisseria gonorrhoeae* using fluorescent amplified fragment length polymorphism analysis. J Clin Microbiol. 2001;39(6):2325–9.

137. Goulding JN, Hookey J V, Stanley J, Olver W, Neal KR, Ala'Aldeen DA, et al. Fluorescent amplified-fragment length polymorphism genotyping of *Neisseria meningitidis* identifies clones associated with invasive disease. J Clin Microbiol. 2000;38(12):4580–5.

138. Patel D, Ellington MJ, Hope R, Reynolds R, Arnold C, Desai M. Identification of genetic variation exclusive to specific lineages associated with *Staphylococcus aureus* bacteraemia. J Hosp Infect. 2015;91(2):136–45.

139. Valicente FH, da Silva RB. fAFLP analysis of Brazilian *Bacillus thuringiensis* isolates. Springerplus. 2014;3:256.

140. Chakrabarti A, Rudramurthy SM, Kale P, Hariprasath P, Dhaliwal M, Singhi S, et al. Epidemiological study of a large cluster of fungaemia cases due to *Kodamaea ohmeri* in an Indian tertiary care centre. Clin Microbiol Infect. 2014;20(2):O83-9.

141. Prashanth K, Pasricha G, Sharma S. Fluorescence amplified fragment length polymorphism for subtyping of genotypes of Acanthamoeba isolated from patients with keratitis. Indian J Med Res. 2011;133:83–7.

142. Kober MV, Abreu MB, Bogo MR, Ferreira CAS, Oliveira SD. Differentiation of *Salmonella enteritidis* isolates by fluorescent amplified fragment length polymorphism. Foodborne Pathog Dis. 2011;8(1):19–26.

143. Slack AT, Symonds ML, Dohnt MF, Craig SB, Smythe LD. Molecular epidemiology of *Leptospira borgpetersenii* serovar Arborea, Queensland, Australia, 1998-2005. Am J Trop Med Hyg. 2010;83(4):820–1.

144. Prashanth K, Singh SK, Kanungo R, Sharma S, Shashikala P, Joshi S, et al. Correlations between genotyping and antibiograms of clinical isolates of *Pseudomonas aeruginosa* from three different south Indian hospitals. Indian J Med Microbiol. 2010;28(2):130–7.

145. Fry NK, Afshar B, Visca P, Jonas D, Duncan J, Nebuloso E, et al. Assessment of fluorescent amplified fragment length polymorphism analysis for epidemiological genotyping of *Legionella pneumophila*

serogroup 1. Clin Microbiol Infect. 2005;11(9):704–12.

146. Thorne N, Evans JT, Smith EG, Hawkey PM, Gharbia S, Arnold C. An IS*6110*-targeting fluorescent amplified fragment length polymorphism alternative to IS*6110* restriction fragment length polymorphism analysis for *Mycobacterium tuberculosis* DNA fingerprinting. Clin Microbiol Infect. 2007b;13:964–70.

147. Borrell S, Thorne N, Espanol M, Mortimer C, Orcau A, Coll P, et al. Comparison of four-colour IS*6110*-fAFLP with the classic IS*6110*-RFLP on the ability to detect recent transmission in the city of Barcelona , Spain. Tuberculosis (Edinb). 2009;89:233–7.

148. Desikan S, Narayanan S. Genetic Markers , Genotyping Methods & Next Generation Sequencing in *Mycobacterium Tuberculosis*. Indian J Med Res. 2015;141(June):761–74.

149. Mardis ER. Next-generation DNA sequencing methods. Annu Rev Genomics Hum Genet. 2008;9:387–402.

150. Singh P, Cole ST. *Mycobacterium leprae*: genes, pseudogenes and genetic diversity. Future Microbiol. 2011;6(1):57–71.

151. Ripoll F, Pasek S, Schenowitz C, Dossat C, Barbe V, Rottman M, et al. Non mycobacterial virulence genes in the genome of the emerging pathogen *Mycobacterium abscessus*. PLoS One. 2009;4(6):e5660.

152. Supply P, Warren RM, Bañuls A-L, Lesjean S, Van Der Spuy GD,

Lewis L-A, et al. Linkage disequilibrium between minisatellite loci supports clonal evolution of *Mycobacterium tuberculosis* in a high tuberculosis incidence area. Mol Microbiol. 2003;47(2):529–38.

153. Sreevatsan S, Xi P, Stockbauer, Kathryn E. Connell ND, Kreiswirth BN, Whittam TS, Musser JM. Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. Proc Natl Acad Sci U S A. 1997;94(September):9869–74.

154. Slatkin M. Linkage disequilibrium--understanding the evolutionary past and mapping the medical future. Nat Rev Genet. 2008;9(6):477–85.

155. Gutacker MM, Smoot JC, Migliaccio CAL, Ricklefs SM, Hua S, Cousins D V, et al. Genome-Wide Analysis of Synonymous Single Nucleotide Polymorphisms in *Mycobacterium tuberculosis* Complex Organisms : Resolution of Genetic Relationships Among Closely Related Microbial Strains. Genetics. 2002;1543:1533–43.

156. Ebrahimi-Rad M, Bifani P, Martin C, Kremer K, Samper S, Rauzier J, et al. Mutations in putative mutator genes of *Mycobacterium tuberculosis* strains of the W-Beijing family. Emerg Infect Dis. 2003;9(7):838–45.

157. Baker L, Brown T, Maiden MC, Drobniewski F. Silent nucleotide polymorphisms and a phylogeny for *Mycobacterium tuberculosis*. Emerg Infect Dis. 2004;10(9):1568–77.

158. 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 U S A. 2006;103(8):2869–73.

159. Hershberg R, Lipatov M, Small PM, Sheffer H, Niemann S, Homolka S, et al. High functional diversity in *Mycobacterium tuberculosis* driven by genetic drift and human demography. PLoS Biol. 2008;6(12):e311.

160. Djelouadji Z, Raoult D, Drancourt M. Palaeogenomics of *Mycobacterium tuberculosis*: epidemic bursts with a degrading genome. Lancet Infect Dis. 2011;11(8):641–50.

161. Gagneux S. Host-pathogen coevolution in human tuberculosis. Philos Trans R Soc Lond B Biol Sci. 2012;367(1590):850–9.

162. Arnold C, Thorne N, Underwood A, Baster K, Gharbia S. Evolution of short sequence repeats in *Mycobacterium tuberculosis*. FEMS Microbiol Lett. 2006;256:340–6.

163. Singh UB, Suresh N, Bhanu NV, Arora J, Pant H, Sinha S, et al. Predominant tuberculosis spoligotypes, Delhi, India. Emerg Infect Dis. 2004;10(6):1138–42.

164. Dong H, Shi L, Zhao X, Sang B, Lv B, Liu Z, et al. Genetic diversity of *Mycobacterium tuberculosis* isolates from Tibetans in Tibet, China. PLoS One. 2012;7(3):e33904.

165. Reed MB, Pichler VK, McIntosh F, Mattia A, Fallow A, Masala S, et al. Major *Mycobacterium tuberculosis* lineages associate with patient country of origin. J Clin Microbiol. 2009;47(4):1119–28.

166. Comas I, Chakravartti J, Small PM, Galagan J, Niemann S, Kremer K, et al. Human T cell epitopes of *Mycobacterium tuberculosis* are evolutionarily hyperconserved. Nat Genet. 2010;42(6):498–503.

167. Gutierrez MC, Brisse S, Brosch R, Fabre M, Omaïs B, Marmiesse M, et al. Ancient Origin and Gene Mosaicism of the Progenitor of *Mycobacterium tuberculosis*. PLoS Pathog. 2005;1(1):e5.

168. Yang Z, Yang D, Kong Y, Zhang L, Marrs CF, Foxman B, et al. Clinical relevance of *Mycobacterium tuberculosis plcD* gene mutations. Am J Respir Crit Care Med. 2005;171(12):1436–42.

169. Warren RM, Sampson SL, Richardson M, Van Der Spuy GD, Lombard CJ, Victor TC, et al. Mapping of IS*6110* flanking regions in clinical isolates of *Mycobacterium tuberculosis* demonstrates genome plasticity. Mol Microbiol. 2000;37(6):1405–16.

170. Parsons LM, Somoskövi A, Gutierrez C, Lee E, Paramasivan CN, Abimiku A, et al. Laboratory diagnosis of tuberculosis in resource-poor countries: challenges and opportunities. Clin Microbiol Rev. 2011;24(2):314–50.

171. Sulis G, Roggi A, Matteelli A, Raviglione MC. Tuberculosis:

epidemiology and control. Mediterr J Hematol Infect Dis. 2014;6(1):e2014070.

172. WHO. Global Tuberculosis Report 2015. 2015;XXXIII(2):81–7.

173. NTCP. National Tuberculosis Control Programme Nepal, Annual Report 2015. 2015.

174. Nepal B. Population mobility and spread of HIV across the Indo-Nepal border. J Health Popul Nutr. 2007;25(3):267–77.

175. Hill EB, Wayne LG, Gross WM. Purification of mycobacterial deoxyribonucleic acid. J Bacteriol. 1972;112(3):1033–9.

176. Gong X, Li J, Zhang Y, Hou S, Qu P, Yang Z, et al. Molecular typing of *Legionella pneumophila* from air-conditioning cooling waters using mip gene, SBT, and FAFLP methods. J Microbiol Methods. 2017;139:1–7.

177. De Bustos A, Cuadrado A, Jouve N, Cuadrado A, Cuadrado A. Sequencing of long stretches of repetitive DNA. Sci Rep. 2016;6(1):36665.

178. Kremer K, van Soolingen D, Frothingham R, Haas WH, Hermans PW, Martín C, et al. Comparison of methods based on different molecular epidemiological markers for typing of *Mycobacterium tuberculosis* complex strains: interlaboratory study of discriminatory power and reproducibility. J Clin Microbiol. 1999;37(8):2607–18.

179. Moganeradj K, Abubakar I, McHugh TD, Sonnenberg P, Arnold C. Insertion site mapping for repeated elements in *Mycobacterium tuberculosis*. J Microbiol Methods. 2013;92:192–6.

180. Thiaucourt F, Manso-Silvan L, Salah W, Barbe V, Vacherie B, Jacob D, et al. *Mycoplasma mycoides*, from “mycoides Small Colony” to “capri”. A microevolutionary perspective. BMC Genomics. 2011;12(1):114.

181. Butler WR, Haas WH. Automated DNA Fingerprinting Analysis of *Mycobacterium tuberculosis* Using Fluorescent Detection of PCR Products. Microbiology. 1996;34(7):1801–3.

182. Millán-Lou MI, López-Calleja AI, Colmenarejo C, Lezcano MA, Vitoria MA, del Portillo P, et al. Global study of IS*6110* in a successful *Mycobacterium tuberculosis* strain: clues for deciphering its behavior and for its rapid detection. J Clin Microbiol. 2013;51(11):3631–7.

183. Alonso H, Samper S, Martín C, Otal I. Mapping IS*6110* in high-copy number *Mycobacterium tuberculosis* strains shows specific insertion points in the Beijing genotype. BMC Genomics. 2013;14:422.

184. Gagneux S, Small PM. Global phylogeography of *Mycobacterium tuberculosis* and implications for tuberculosis product development. Lancet Infect Dis. 2007;7(5):328–37.

185. Hanekom M, Gey van Pittius NC, McEvoy C, Victor TC, Van Helden PD, Warren RM. *Mycobacterium tuberculosis* Beijing genotype: a

template for success. Tuberculosis (Edinb). 2011;91(6):510–23.

186. Hanekom M, van der Spuy GD, Streicher E, Ndabambi SL, McEvoy CRE, Kidd M, et al. A Recently Evolved Sublineage of the *Mycobacterium tuberculosis* Beijing Strain Family Is Associated with an Increased Ability to Spread and Cause Disease. J Clin Microbiol. 2007;45(5):1483–90.

187. Philipp WJ, Poulet S, Eiglmeier K, Pascopella L, Balasubramanian V, Heym B, et al. An integrated map of the genome of the tubercle bacillus, *Mycobacterium tuberculosis* H37Rv, and comparison with *Mycobacterium leprae*. Proc Natl Acad Sci U S A. 1996;93(7):3132–7.

188. Moganeradj K, Rajendram D, Khadge S, Sonnenberg P, Abubakar I, D. McHugh T, et al. Insertion Element I*S6110* based characterisation of Nepalese tuberculosis strains into different genetic lineages. Clin Microbiol Infect Dis. 2016;1(1):1–5.

189. 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. 2017;105:60–72.

190. 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.

191. Soto SM. Human migration and infectious diseases. Clin Microbiol Infect. 2009;15(s1):26–8.

192. Pang Y, Zhou Y, Zhao B, Liu G, Jiang G, Xia H, et al. Spoligotyping and drug resistance analysis of *Mycobacterium tuberculosis* strains from national survey in China. PLoS One. 2012;7(3):e32976.

193. Chen H, He L, Huang H, Shi C, Ni X, Dai G, et al. *Mycobacterium tuberculosis* Lineage Distribution in Xinjiang and Gansu Provinces, China. Sci Rep. 2017;7(1):1068.

194. Hameed S, Moganeradj K, Mahmood N, McHugh TD, Chaudhry MN, Arnold C. Sequence analysis of the rifampicin resistance determining region (RRDR) of *rpoB* gene in multidrug resistance confirmed and newly diagnosed tuberculosis patients of Punjab, Pakistan. PLoS One. 2017;12(8).

195. Malla B, Stucki D, Borrell S, Feldmann J, Maharjan B, Shrestha B, et al. First insights into the phylogenetic diversity of *Mycobacterium tuberculosis* in Nepal. PLoS One. 2012;7(12):e52297.

196. Somoskovi A, Parsons LM, Salfinger M. The molecular basis of resistance to isoniazid, rifampin, and pyrazinamide in *Mycobacterium tuberculosis*. Respir Res. 2001;2(3):164–8.

197. Arnold C, Westland L, Mowat G, Underwood A, Magee J, Gharbia S. Single-nucleotide polymorphism-based differentiation and drug

resistance detection in *Mycobacterium tuberculosis* from isolates or directly from sputum. Clin Microbiol Infect. 2005;11(2):122–30.

198. Creswell J, Rai B, Wali R, Sudrungrot S, Adhikari LM, Pant R, et al. Introducing new tuberculosis diagnostics: the impact of Xpert(®) MTB/RIF testing on case notifications in Nepal. Int J Tuberc lung Dis. 2015;19(5):545–51.

199. Alonso H, Aguilo JI, Samper S, Caminero JA, Campos-Herrero MI, Gicquel B, et al. Deciphering the role of IS*6110* in a highly transmissible *Mycobacterium tuberculosis* Beijing strain, GC1237. Tuberculosis (Edinb). 2011;91(2):117–26.

200. Soto CY, Menéndez MC, Pérez E, Samper S, Gómez AB, García MJ, et al. IS*6110* mediates increased transcription of the *phoP* virulence gene in a multidrug-resistant clinical isolate responsible for tuberculosis outbreaks. J Clin Microbiol. 2004;42(1):212–9.

201. Lee CE, Goodfellow C, Javid-Majd F, Baker EN, Shaun Lott J. The crystal structure of *TrpD*, a metabolic enzyme essential for lung colonization by *Mycobacterium tuberculosis*, in complex with its substrate phosphoribosylpyrophosphate. J Mol Biol. 2006;355(4):784–97.

202. Sreejit G, Ahmed A, Parveen N, Jha V, Valluri VL, Ghosh S, et al. The ESAT-6 protein of *Mycobacterium tuberculosis* interacts with beta-2-microglobulin (2M) affecting antigen presentation function of

macrophage. PLoS Pathog. 2014;10(10):e1004446.

203. Guo S, Xue R, Li Y, Wang SM, Ren L, Xu JJ. The CFP10/ESAT6 complex of *Mycobacterium tuberculosis* may function as a regulator of macrophage cell death at different stages of tuberculosis infection. Med Hypotheses. 2012 Mar;78(3):389–92.

204. Chen Y-Y, Chang J-R, Huang W-F, Kuo S-C, Su I-J, Sun J-R, et al. Genetic diversity of the *Mycobacterium tuberculosis* Beijing family based on SNP and VNTR typing profiles in Asian countries. PLoS One. 2012;7(7):e39792.

205. Mokrousov I, Ly HM, Otten T, Lan NN, Vyshnevskyi B, Hoffner S, et al. Origin and primary dispersal of the *Mycobacterium tuberculosis* Beijing genotype: clues from human phylogeography. Genome Res. 2005;15(10):1357–64.

206. Siddiqi N, Shamim M, Hussain S, Choudhary RK, Ahmed N, Prachee, et al. Molecular characterization of multidrug-resistant isolates of *Mycobacterium tuberculosis* from patients in North India. Antimicrob Agents Chemother. 2002;46(2):443–50.

207. Oliveira CLN, Muniz-Sobrinho J da S, Viana-Magno LA, Melo OSC, Macho A, Rios-Santos F. Detection of multidrug-resistant *Mycobacterium tuberculosis* strains isolated in Brazil using a multimarker genetic assay for *katG* and *rpoB* genes. Braz J Infect Dis. 2016;20(2):166–72.

208. Tripathi R, Sinha P, Kumari R, Chaubey P, Pandey A, Anupurba S. Detection of rifampicin resistance in tuberculosis by molecular methods: A report from Eastern Uttar Pradesh, India. Indian J Med Microbiol. 2016;34(1):92–4.

209. Pantoja A, Kik S V., Denkinger CM. Costs of Novel Tuberculosis Diagnostics--Will Countries Be Able to Afford It? J Infect Dis. 2015;211(suppl 2):S67–77.

Appendices

Appendix 1. Sequences (genes of interest, see Table 4-1) using the DNASTAR Lasergene8 software where the IS*6110* insertion flanks the genome

1. IS*6110* Sequence orientation *per se* in the whole genome:
a) 1543249 to 1543307(58bp) to next *TaqI* site
CCGCGCCGGCTGGCAACCGTTCCCGCT (27bp) ~102bp
b) 1989000 to 1989058 (58bp) to next *TaqI* site
AGAGGACTTCAT (12bp) ~87bp
c) 2431414 to 2431472 (58bp) to next *TaqI* site
GGGCTTCCCGAGACTGCGATTCCCAAACGATGACGCCCAAACAAAAAGCGGGA CCGCCGATGGCTGCCCCGCTGCCGCTGGTTGCGTTCGGCTTACTCGT (100bp) ~ 175bp
d) 3121821- 3121879 (58bp) to next *TaqI* site
GTTTTGGGTCTGACGACTCGCGGCGAGCACGTCTCACCCAGCAGGCGGTGAG GTTGGGTTTCCGTCCCCTCTCGGGGTTTTGGGTCTGACGACACGGACGAGCTG GACCGCATCAGCGATGCTGAGCTGAGGGTTTCCGTCCCCTCTCGGGGTTTTGG GTCTGACGACTTGTCTCAATCGTGCCGTCTGCGGTGACACGCTCCAAGTTTCC GTCCCCTCTCGGGGTTTTGGGTCTGACGACCACCAGGATCAGCGCCAAGCCA GTTAGCGCAATCCAGTTTCCGTCCCCTCTCGGGGTTTTGGGTCTGACGACCTC CCGGACCATCTGCAGCTCGCCCGGGTCCATGCGGTTTCCGTCCCCTCTCGGG GTTTTGGGTCTGACGACCGGAGTCATCCGCGCGGGCCGGCGCGATTGTTGCC GGGTTTCCGTCCCCTCTCGGGGTTTTGGGTCTGACGACTGGCGATTTACGACG CTGACGGGAACTCGTGCGAATGTTTCCGTCCCCTCTCGGGGTTTTGGGTCTGA TCCGCGAAATTCACTGCGCGTTATTCAAGGTTTCCGTCCCCTCTCGGGGTTTTG

GGTCTGACGACCCGAGCCGACCATCCGCATCACACCGAAAGGTTGGCGCAA
GTTTCCGTCCCCTCTCGGGGTTTTGGGTCTGACGACACGTGGGGAGAGGGAAT
GGCAATGATGGT (697bp) ~772bp

2. **Reverse Complement of the whole genome and the IS*6110* Sequence orientation *per se*:**

a) 520697 to 520755 (58bp) to next *TaqI* site

TCCGTATCGTCGCCGACCGCTTGGGAGACCCGCGAGCGCGTGCCGTGGTGTT
CTTGCATGGCGGTGGACAGACGCGACGCTCATGGGGTCGGGCCGCAGCCGCT
GTCGCCGAACGTGGCTGGCAAGCGGTCACGAT (136bp~201bp

b) 701094 to 701152 (58bp) to next *TaqI* site

TGTCCATAACAAGCGGTCGGTGCGGTTGACTTGTTGGACGCGCCAGATGCACT
GTTTGACTCGCGT (66bp) ~141bp

c) 1775899 to 1775957 (58bp) to next *TaqI* site

AAACTTCGGTAGCCAAAACCTGGGTAGCGGCAACATCGGCAGCACCAACGTGG
GCAGCGGCAACATCGGCAGCACCAACGTGGGCAGCGGCAACATCGGCGACAC
GAACTTCGGTAACGGAAACAACGGCAACTTCAACTTTGGTAGCGGCAATACCG
GCAGTAACAACATCGGCTTCGGAAACACCGGCAGCGGGAATTTCGGTTTCGGA
AACACGGGCAACAACAACATCGGTATCGGGCTCACCGGCGATGGTCAGATCG
GCATCGGCGGACTGAACTCGGGCAGCGGAAACATCGGTTTCGGGAACTCCGG
CACCGGAAACGTCGGTTTGTTCAACTCCGGCACCGGCAACGTAGGCTTCGGGA
ACTCCGGTACTGCGAACACTGGATTCGGGAACGCGGGCAACGTCAACACCGG
ATTTTGGAACGGCGGCAGCACAAACACTGGCCTCGCTAACGCCGGCGCCGGC
AACACAGGCTTTTT (486bp) ~561bp

d) 2415375 to 2415433 (58bp) to next *TaqI* site

GTTGCCGACCTTTGTCCTGTTGATCCTGTTGGCGGCGGTGACTGCCGTCTACG

CACTACGCGGCGCGTTACGCGCTGCCACCTCACTCATGCGCGGCCGGCGCGG
TGCCGACCGGTAGCGT (121bp) ~196bp

3. IS*6110* Sequence orientation *per se* in the whole genome (terminal G replaced)

a) 2785912 to 2785969 (58bp) to next *TaqI* site

Too short to be seen in vitro

b) 3796355 to 3796412 (58bp) to next *TaqI* site

AATCAAGTCCCCGCGTCCGTTGCGAATCGTGGTTGTCATTGCGCGCGAACCTG
TTTGGGAAGGCCGAATCGCACCGTCTCGGTCGCTAT (89bp) ~164bp

4. Reverse Complement of the whole genome and the IS*6110* Sequence orientation *per se* (terminal G replaced)

a) 858763 to 858820 (58 bases) to next *TaqI* site

AATCCGTACAACCGCGACGCGGCGGAACAGGCTGGACGGCGGAATAGAGGCT
GATCTGGACGTATGAGACCGACCGCGATGGGAGAGATTGCGACTGGTTAGCTC
ATAACCCAGAGGTCGCAGGTTCGTGAACCGCCCCGGTGAGTCCGGAGACTCT
CTGATCTGAGACCTCAGCCGGCGGCTGGTCTCTGGCGTTGAGCGTAGTAGGC
AGCCT (214bp) ~289bp

b) 1439367 to 1439424 (58bp) to next *TaqI* site

GAAGATGAGCTGCGCAAGCGTGGACGTGAGCGTCGCCCGTTGGACAAGTTGC
CGCCGGGGCCCATCCCGGGCACCGGCGGTCAGCCGCTCCAACCGTTTAAACC
ATCCCGGTAGCCAGCTTCGCTGCGGGTTGCGTTGT (139bp) ~214bp

c) 1861462 to 1861519 (58 bases) to next *TaqI* site

GGTGGCGCTGGTGATCGCCTTGGGAGGGACTTGCGGTGTGGCGGATGCTTTG
CCGCTGGGCCAGACTGACGACCCGATGATTGTCGCGCATCGCGCGGGTACGC

GCGATTTTCCCGAGAACACCGTTCTGGCGATTACCAACGCAGTCGCAGCTGGTGTGGATGGCATGTGGCTGACCGTCCAGGTCAGCAGCGATGGCGTGCCGGTGCTGTATCGTCCGTCCGATCTGGCAACGTTGACCGACGGCGCCGGCCCGGTGAATT (263bp) ~338bp

d) 2046062 to 2046119 (58 bases) to next *TaqI* site

CGCGCGGCTACAAAAAAGCTTGTATCACTAGCTAGAGTGGAGTCGGTTTGCT (52bp) ~127bp

e) 3522455 to 3522512 (58 bases) to next *TaqI* site

CCTCCTTCACTTGAGGGGGTGTTGAGGTTGGGCCGGTCGGCGGACACATCTCAGTCGGGGCGAAAGCCACGCTCCTATCAAGTCACGCCGGCCGGTCCTTCTCACCTAGTGCCGGCAATACTGCTCAACGCCAACCCCAAGAGGACGGCAGACGCAAAGTGAGCCAGACACCAGGTGATCCGGAACAAACCACCGCGACACGGCGTCTGTCACACCGACACACCCACCTCGCGGCACACACAACACCGACACTGAGACATAAAGGTCCCTTTTCGCACGGCGTGT (285bp) ~360bp

Appendix 2: Nepal strains

N1-N176

Each strain has a list of fragments that are generated from differing insertion points of the IS*6110* element. Each fragment is listed as a size in base pairs and can be one of four colours, R = red, B = blue, G = green and Y = yellow. For example 80.4Y is an 80.0 base pair fragment which is yellow.

If two strains share an identical IS*6110* insertion site then they will each possess a fragment of the EXACT same size and colour. For example, below N36 and N34 share most of the same fragments, including: 81.0B, 100.0G, 124.9G, 140.4R, 197.2B, 207.1Y, 264.2G, 360.0R, 401.5B and 560.4B. The majority of these fragments are also shared by H37Rv and H37Ra but not all (i.e. they are not identical but very similar).

Unassigned

N102

84.4R, 106.6G, 312.8G

N107

84.4R, 106.6G, 312.8G

N104

84.4R, 106.6G, 312.8G

N25

178.1R, 237.8 G, 300.8G

N62

74.9Y, 78.2R

PGG3 T group (H37Rv-like)

N108

88.9Y, 106.2G, 179.1B, 183.6R, 254.3Y, 301.5G, 312.2G, 360.0R, 443.1R

N36

80.4Y, 81.0B, 83.8Y, 100.0G, 124.9G, 140.4R, 193.2R, 197.2B, 207.1Y, 264.2G, 360.0R, 401.5B, 560.4B

N34

81.0B, 100.0G, 124.9G, 139.1R, 140.4R, 192.4R, 197.2B, 207.1Y, 264.2G, 360.0R, 401.5B, 560.4B

N163

81.0B, 83.61Y, 99.4G, 119.54G, 137.58Y, 139.91R, 193.44R, 197.6B, 207.14Y, 264.24G, 312.29G, 360.71R, 561.33B

CAS group

N103

78.4G, 88.2R, 92.0B, 93.5B, 114.3Y, 127.7R, 148.4R, 207.1G, 249.6G, 275.1R, 394.7R, 434.3R

N101

74.4G, 92.0B, 115.6B, 116.8G, 123.9B, 127.7R, 136.0Y, 157.8G, 275.1R, 316.4R, 432.9R

N116

106.6G, 116.8G, 117.9R, 127.7R, 206.5G, 244.7R, 275.1R, 286.9G, 394.7R, 433.6R

N109

92.0B, 116.1Y, 117.9R, 171.9R, 195.8R, 206.2G, 263.2B, 268.3R, 275.1R, 341.4G, 434.3R, 461.1G

N111

84.4R, 106.6G, 121.6B, 142.0B, 142.5Y, 145.1R, 147.6B, 159.5Y, 186.4B, 194.2R, 240.9Y, 285.6B, 312.2G, 326.4Y

N113

77.3R, 83.5G, 87.0Y, 96.7B, 97.8G, 100.8R, 105.5R, 112.4G, 124.0R, 127.1R 133.7R, 169.7Y, 176.8Y, 220.5B, 320.5R, 376.5R, 432.9R

N114

84.9B, 102.1B, 102.5Y, 112.4G, 127.1R, 139.5R, 158.5Y, 181.2Y, 248.3R, 255.3G, 334.9R, 354.3B, 461.1G

N12

74.4G, 92.0B, 111.4G, 117.9R, 127.7R, 131.2Y, 170.7R, 206.2G, 232.3R, 275.1R, 393.0R, 432.9R, 560.0G

N50

78.4G, 92.0B, 117.9R, 127.7R, 131.2Y, 178.1R, 206.2G, 275.1R, 321.6R, 393.0R, 432.9R, 560.0G

N73

77.3G, 86.1G, 92.0B, 117.9R, 127.7R, 206.2G, 275.1R, 393.0R, 432.9R, 493.1G,

N36

78.4G, 83.3R, 92.0B, 127.7R, 188.9Y, 192.8R, 198.3B, 206.2G, 275.1R, 324.7G, 393.0R, 432.9R,

N37

78.4G, 92.0B, 117.9R, 127.7R, 206.2G, 275.1R, 393.0R, 432.9R,

N67

78.4G, 92.0B, 117.9R, 127.7R, 142.6R, 150.8R, 206.2G, 275.1R, 316.4R, 318.7Y, 393.0R, 432.9R, 493.1G,

N68

77.3G, 92.0B, 117.9R, 127.7R, 142.6R, 150.8R, 206.2G, 275.1R, 316.4R, 318.7Y, 393.0R, 432.9R,

N7

77.3G, 92.0B, 106.6G, 117.9R, 127.7R, 131.1G, 150.8R, 206.2G, 247.3R, 275.1R, 316.4R, 393.0R, 432.9R,

N18

74.4G, 78.4G, 92.0B, 127.7R, 150.8R, 157.8G, 178.1R, 178.2Y, 206.2G, 229.5R, 275.1R, 316.4R, 321.6R, 393.0R, 432.9R, 778B

N17

78.4G, 92.0B, 115.6B, 116.8G, 127.7R, 136.0Y, 157.8G, 275.1R, 316.4R, 393.0R, 432.9R,

N10

74.4G, 92.0B, 127.7R, 149.4Y, 150.8R, 206.2G, 230.1Y, 247.3R, 275.1R, 316.4R, 379.7R, 393.0R, 432.9R,

N72

78.4G, 92.0B, 127.7R, 149.4Y, 150.8R, 206.2G, 230.1Y, 247.3R, 275.1R, 316.4R, 379.7R, 393.0R, 432.9R,

N19

78.4G, 92.0B, 96.2R, 114.5B, 127.7R, 136.0Y, 148.0Y, 150.8R, 171.9R, 206.2G, 275.1R, 316.4R, 393.0R, 432.9R,

N6

77.3G, 88.2R, 92.0B, 93.5B, 114.3Y, 117.9R, 127.7R, 148.4R, 206.2G, 275.1R, 328.4G, 393.0R, 432.9R, 455.9G,

N3

77.3G, 88.2R, 92.0B, 93.5B, 114.3Y, 117.9R, 127.7R, 148.4R, 206.2G, 275.1R, 393.0R, 455.9G,

N91

78.4G, 88.2R, 92.0B, 93.5B, 114.3Y, 117.9R, 127.7R, 138.9G, 206.2G,299.2Y,

N5

78.4G, 92.7B, 101.5Y, 107.1R, 116.7R, 117.9R, 183.8G, 206.2G, 251.8R, 275.1R, 321.6R, 393.0R, 432.9R,

N56

92.0B, 117.9R, 160.3Y, 206.2G, 318.0B, 352.9G, 393.0R, 432.9R,

N65

77.3G, 92.0B, 101.2B, 129.8G, 132.0R, 305.9R, 352.9G, 393.0R, 432.9R,

N29

78.4G, 81.0B, 92.0B, 100.5Y, 117.9R, 127.7R, 194.2R, 354.2Y,

N69

78.4G, 81.0B, 92.0B, 100.5Y, 117.9R, 127.7R, 194.2R, 266.1G, 275.1R, 310.6B, 316.5B, 353.7Y, 393.0R, 494.3B,

N48

94.8B, 117.9R, 275.1R, 296.9Y, 297.4R, 302.4G, 393.0R, 432.9R, 437.5B,

N58

94.8B, 117.9R, 275.1R, 296.9Y, 297.4R, 302.4G, 393.0R, 432.9R, 437.5B,

N54

78.4G, 81.7Y, 91.0R, 92.0B, 107.5Y, 112.4G, 127.7R, 142.6R, 150.8R, 216.3R, 318.7Y, 393.0R,

N81

78.4G, 81.7Y, 91.0R, 92.0B, 107.5Y, 114.3Y, 127.7R, 142.6R, 150.8R, 318.7Y, 345.5R,

N28

78.4G, 92.0B, 129.8G, 136.1G, 206.2G, 275.1R, 283.0R, 303.3G, 354.0G, 393.0R, 432.9R, 565.5R,

N93

78.4G, 92.0B, 122.8B, 136.1G, 176.8Y, 185.3G, 187.0Y, 275.1R, 393.0R, 432.9R, 563.6R

N26

73.4Y, 78.4G, 79.0Y, 79.4G, 87.4R, 127.7R, 206.2G, 239.3G, 256.2B, 275.1R, 321.6R, 393.0R, 432.9R, 445.3G, 868Y

N32

78.4G, 79.0Y, 92.0B, 127.7R, 206.2G, 239.3G, 256.2B, 273.1Y, 275.1R, 321.6R, 393.0R, 432.9R, 868Y

N45

87.3Y, 92.0B, 93.0G, 94.5R, 118.0G, 127.7R, 156.0Y, 206.5R, 279.5Y, 393.0R, 432.9R,

N75

92.0B, 115.6B, 116.8G, 127.7R, 135.9R, 151.4R, 275.1R, 317.0R, 370.5G, 393.0R, 432.9R, 461.1G,

N61

94.8B, 129.8G, 160.3Y, 206.2G, 275.1R, 318.0B, 353.5G, 393.0R, 432.9R,

N78

91.5G, 100.3R, 120.0R, 132.2Y, 152.4R, 173.2B, 206.2G, 275.1R, 317.0R, 351.9R, 393.0R, 432.9R,

N51

92.0B, 93.0G, 206.5G, 593.2G

N124

78.21G, 84.4R, 93.09G, 96.01R, 124.97Y, 127.59R, 139.5678.21G, 84.4R, 93.09G, 96.01R, 124.97Y, 127.59R, 139.56Y, 206.16G, 256.05Y, 275.31R, 316.82R, 326.61R, 433.6R, 445.51R

N135

78.25G, 83.63Y, 87.15R, 130.04G, 133.82R, 136.22Y, 138.89R, 155.07R, 167.53Y, 206.67G, 261.98B, 275.49R, 354.6G, 394.57R, 434.01R, 450.43R

N138

78.15G, 92.07B, 101.69B, 115.49G, 129.93G, 306.59R, 353.9G, 394.41G, 433.88R

N139

78.07G, 115.02G, 129.43G, 203.71G, 305.58R, 309.28G, 352.67G, 393.1R

N140

78.15G, 88.2Y, 91.65R, 95.98R, 101.58B, 110.94B, 115.58G, 129.99G, 266.22G, 274.13G, 306.57R, 353.84G, 394.28R, 433.81R

N141

84.48R, 86.86Y, 89.24G, 92.51R, 106.18G, 128.76G, 191.13G, 194.09R, 255.86R, 333.21G, 431.28R

N143

78.24G, 92.13B, 101.68B, 124.35G, 129.99G, 306.74R, 331.35G, 354.06G, 394.66R, 434.09R

N144

78.14G, 93.22G, 101.69B, 130.09G, 144.17G, 306.74R, 354.05G, 394.66R, 434R

, N145, N146, N148, N155, N158, N161, N165, N167, N168, N173, N174, N176

Haarlem

N39

87.0Y, 148.7B, 149.6G, 151.4R, 224.3G, 265.6B, 300.2R, 311.0G, 445.7Y,

N87

87.0Y, 116.1R, 148.7B, 151.4R, 224.3G, 265.6B, 300.2R, 311.0G, 445.7Y, 787.0B

N110

87.0Y, 116.1R, 148.7B, 151.4R, 167.7G, 177.5R, 224.3G, 265.6B, 301.1R, 312.2G, 446.6Y, 787.0B

N106

87.3Y, 89.2B, 124.0R, 149.1B, 194.2R, 224.3G, 266.1B, 301.1R, 312.8G, 446.6Y

N137, N152, N153, N157, n160

X group

N27

311.0G, 300.2R

N77

80.4Y, 83.7Y, 301.1R 311.0G

N42

80.4Y, 83.7Y, 148.4R, 301.1R, 410.3R

N129

LAM group

N53

71.5Y, 96.7B, 105.2R, 116.1R, 121.2R, 161.2B, 164.6Y, 338.6Y, 441.9Y

N64

88.6G, 96.7B, 102.4G, 105.2R, 116.1R, 136.9Y, 227.5R, 431.9R, 444.1Y,

N9

71.5Y, 96.7B, 103.0B, 105.2R, 191.3G, 198.6R, 256.5R, 333.9G, 431.9R, 463.2R, 606.6B

N127

Beijing

N115

84.5Y, 95.7Y, 101.7B, 102.5Y, 112.4G, 115.6B, 127.1R, 139.1R, 180.7Y, 195.1Y, 199.5B, 254.8G, 334.9R, 354.3B, 389.8Y, 461.1G

N24

84.5Y, 95.7Y, 101.7B, 102.5Y, 103.0B, 112.4G, 127.1R, 131.2Y, 139.1R, 180.7Y, 234.8G, 254.8G, 265.7G, 309.9R, 332.4R, 353.5B, 389.8Y, 433.6B,

N8

80.0R, 84.5Y, 95.7Y, 101.2B, 102.5Y, 112.4G, 127.1R, 139.1R, 180.3R, 254.8G, 265.1G, 309.9R, 332.4R, 353.5B, 389.8Y, 432.7B, 458.8G,

N23

84.5Y, 95.7Y, 101.7B, 102.5Y, 112.4G, 127.1R, 135.5G, 139.1R, 140.7R, 180.7Y, 254.8G, 332.4R, 353.5B, 389.8Y, 433.6B, 460.3G,

N70

84.5Y, 93.5B, 95.7Y, 101.7B, 102.5Y, 112.4G, 127.1R, 139.1R, 147.7Y, 180.7Y, 201.1G, 254.8G, 266.1G, 309.9R, 332.4R, 353.5B, 389.8Y, 433.6B, 460.3G,

N40

84.5Y, 95.7Y, 101.7B, 102.5Y, 112.4G, 115.5B, 127.1R, 139.1R, 180.7Y, 254.8G, 265.7G, 309.9R, 332.4R, 353.5B, 389.8Y, 433.6B, 460.3G,

N33

84.5Y, 95.7Y, 101.7B, 102.5Y, 112.4G, 127.1R, 139.1R, 180.7Y, 309.9R, 332.4R, 353.5B

N82

84.5Y, 95.7Y, 101.7B, 102.5Y, 112.4G, 127.1R, 135.5G, 139.1R, 140.7R, 180.7Y, 254.8G, 332.4R, 333.5R, 353.5B, 389.8Y, 433.6B, 460.3G,

N35

84.5Y, 95.7Y, 101.7B, 102.5Y, 112.4G, 127.1R, 139.1R, 180.7Y, 246.5B, 254.8G, 266.1G, 309.9R, 332.4R, 353.5B, 389.8Y, 433.6B, 460.3G

N14

84.5Y, 95.7Y, 101.7B, 102.5Y, 107.5Y, 112.4G, 125.2Y, 127.1R, 139.1R, 180.7Y, 254.8G, 332.4R, 353.5B, 390.5Y, 434.5B, 460.3G,

N74

84.5Y, 95.7Y, 101.7B, 112.4G, 127.1R, 139.1R, 180.7Y, 254.8G, 266.1G, 309.9R, 332.4R, 353.5B, 389.8Y, 433.6B, 460.3G,

N76

95.7Y, 101.7B, 112.4G, 114.9B, 127.1R, 139.1R, 181.2Y, 195.1Y, 253.2B, 254.8G, 266.1G, 332.4R, 309.9R, 353.5B, 389.8Y, 433.6B, 460.3G,

N90

95.7Y, 101.7B, 112.4G, 127.1R, 139.1R, 181.2Y, 253.2B, 254.8G, 266.1G, 309.9R, 332.4R, 353.5B, 389.8Y, 433.6B, 460.3G,

N47

95.7Y, 101.7B, 112.4G, 127.1R, 129.8G, 182.8G, 254.8G, 353.5B, 433.6B, 460.3G,

N55

95.7Y, 101.7B, 112.4G, 254.8G, 266.1G, 460.3G,

N59

95.7Y, 101.7B, 112.4G, 127.1R, 223.7G, 246.5B, 254.8G, 266.1G, 353.5B, 460.3G,

N120, N125, N136, N130, N131, N132, N133, N134, N149, N151, N154, N175

M. africanum

N31

87.1B, 88.6G, 134.7Y, 140.4R, 152.4R, 183.6R, 195.1R, 195.8R, 239.3Y, 287.2R, 302.5R, 598.0R

N164

PGG1 Outliers

N13

87.4R, 90.2B, 95.6B, 96.4G, 187.2R, 219.8R, 311.0G, 374.6R, 445.3G,

N20

95.6B, 311.4G

N21

137.1R, 150.8G, 224.6R, 310.5G, 435.4Y

N38

107.3G, 117.8Y, 137.1R, 150.8G, 166.3R

N22

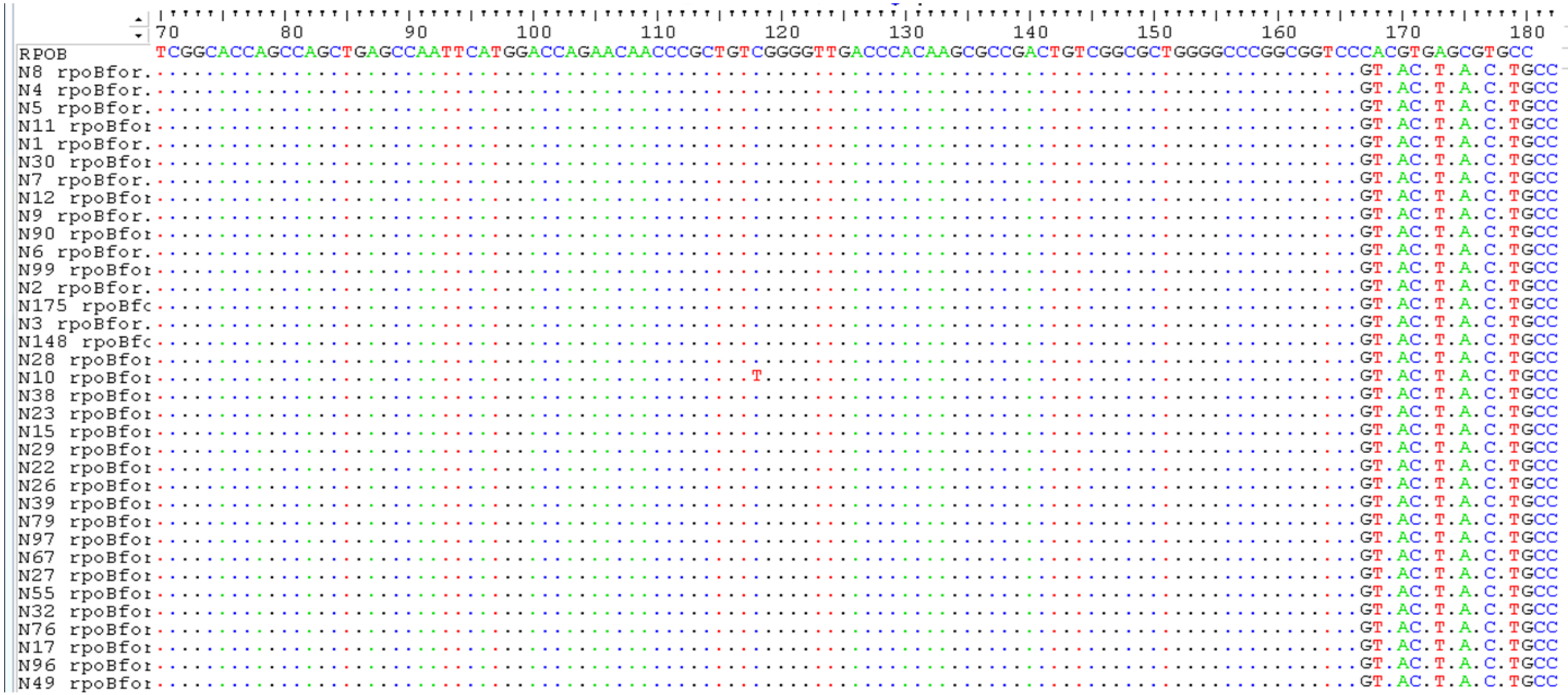
190.3R, 224.6R, 239.9B, 253.4G, 263.2G, 309.9G, 404.3G,

N4

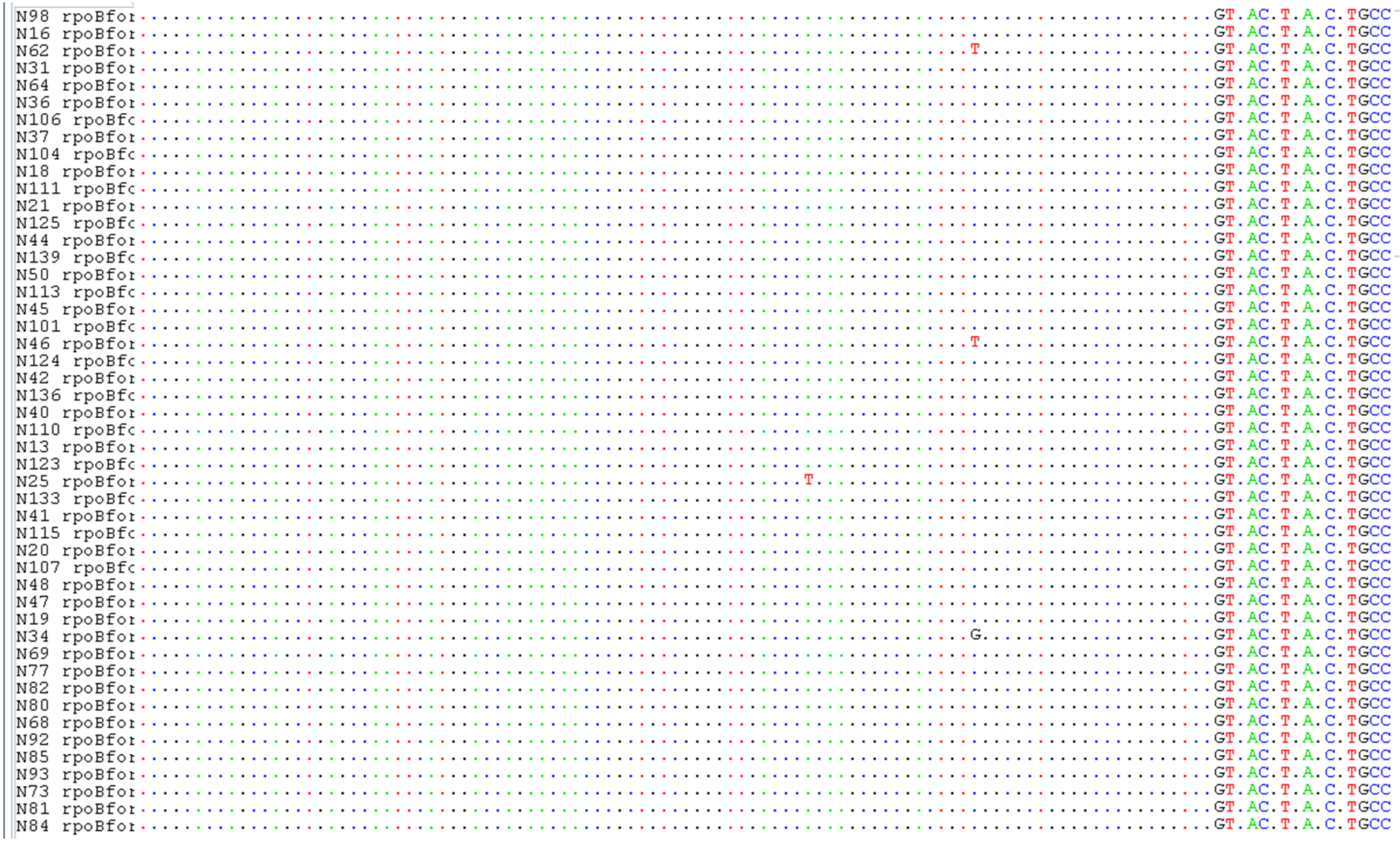
171.9B, 196.1G, 225.8Y, 253.4G, 262.7G, 272.7R, 308.7G, 356.4Y, 403.5G, 505.2Y,

N121, N122, N123, N142, N150, N159, N162, N166, N169, N170, N171, N172

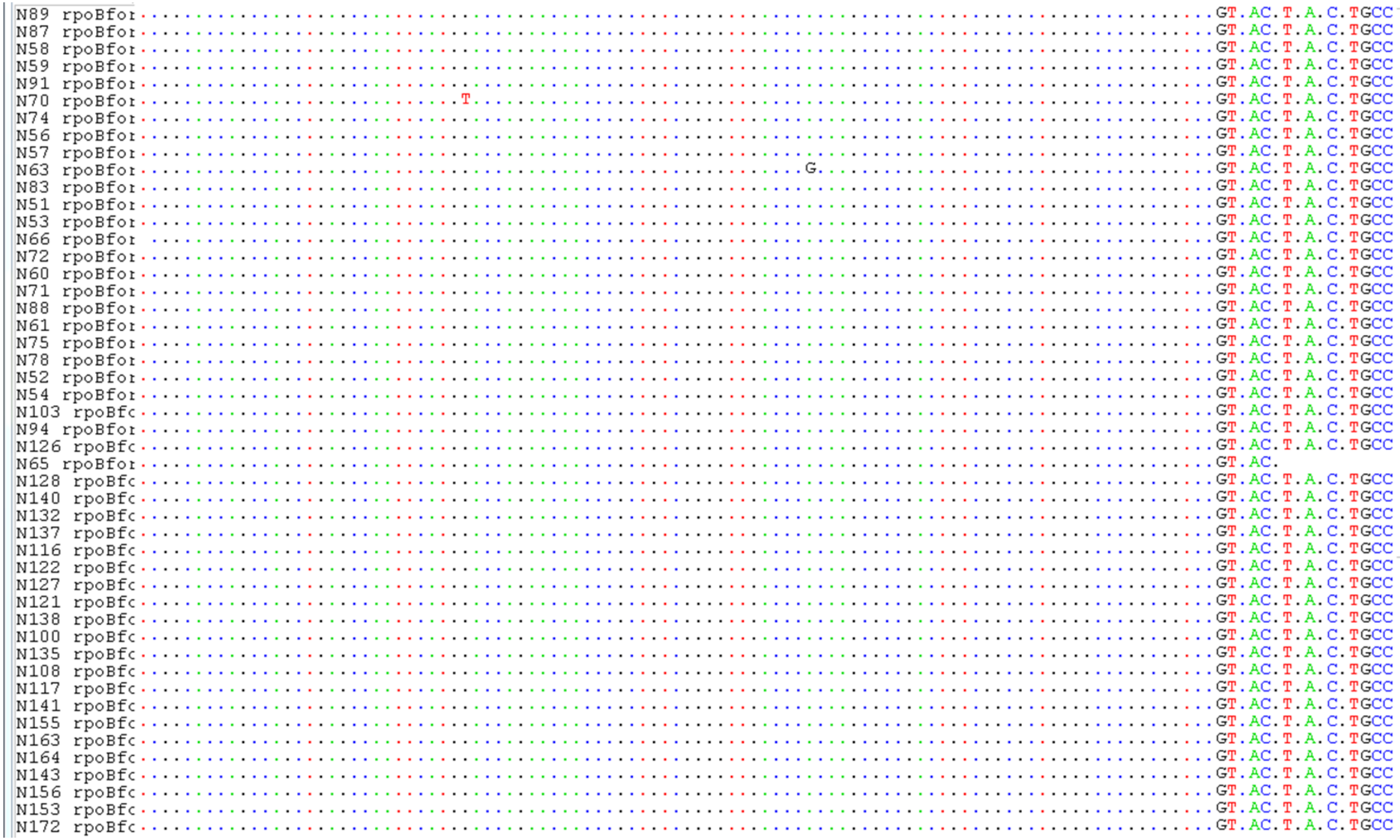
Appendix 3: rpoB mutations seen in the RRDR region using Clustal W alignment in BioEdit sequence alignement software



```
            70        80        90        100       110       120       130       140       150       160       170       180
RPOB        TCGGCACCAGCCAGCTGAGCCAATTCATGGACCAGAACAACCCGCTGTCGGGGTTGACCCACAAGCGCCGACTGTCGGCGCTGGGGCCCGGCGGTCCCACGTGAGCGTGCC
N8 rpoBfor. ................................................................................................GT.AC.T.A.C.TGCC
N4 rpoBfor. ................................................................................................GT.AC.T.A.C.TGCC
N5 rpoBfor. ................................................................................................GT.AC.T.A.C.TGCC
N11 rpoBfor ................................................................................................GT.AC.T.A.C.TGCC
N1 rpoBfor. ................................................................................................GT.AC.T.A.C.TGCC
N30 rpoBfor ................................................................................................GT.AC.T.A.C.TGCC
N7 rpoBfor. ................................................................................................GT.AC.T.A.C.TGCC
N12 rpoBfor ................................................................................................GT.AC.T.A.C.TGCC
N9 rpoBfor. ................................................................................................GT.AC.T.A.C.TGCC
N90 rpoBfor ................................................................................................GT.AC.T.A.C.TGCC
N6 rpoBfor. ................................................................................................GT.AC.T.A.C.TGCC
N99 rpoBfor ................................................................................................GT.AC.T.A.C.TGCC
N2 rpoBfor. ................................................................................................GT.AC.T.A.C.TGCC
N175 rpoBfo ................................................................................................GT.AC.T.A.C.TGCC
N3 rpoBfor. ................................................................................................GT.AC.T.A.C.TGCC
N148 rpoBfo ................................................................................................GT.AC.T.A.C.TGCC
N28 rpoBfor ................................................................................................GT.AC.T.A.C.TGCC
N10 rpoBfor ................................................T...............................................GT.AC.T.A.C.TGCC
N38 rpoBfor ................................................................................................GT.AC.T.A.C.TGCC
N23 rpoBfor ................................................................................................GT.AC.T.A.C.TGCC
N15 rpoBfor ................................................................................................GT.AC.T.A.C.TGCC
N29 rpoBfor ................................................................................................GT.AC.T.A.C.TGCC
N22 rpoBfor ................................................................................................GT.AC.T.A.C.TGCC
N26 rpoBfor ................................................................................................GT.AC.T.A.C.TGCC
N39 rpoBfor ................................................................................................GT.AC.T.A.C.TGCC
N79 rpoBfor ................................................................................................GT.AC.T.A.C.TGCC
N97 rpoBfor ................................................................................................GT.AC.T.A.C.TGCC
N67 rpoBfor ................................................................................................GT.AC.T.A.C.TGCC
N27 rpoBfor ................................................................................................GT.AC.T.A.C.TGCC
N55 rpoBfor ................................................................................................GT.AC.T.A.C.TGCC
N32 rpoBfor ................................................................................................GT.AC.T.A.C.TGCC
N76 rpoBfor ................................................................................................GT.AC.T.A.C.TGCC
N17 rpoBfor ................................................................................................GT.AC.T.A.C.TGCC
N96 rpoBfor ................................................................................................GT.AC.T.A.C.TGCC
N49 rpoBfor ................................................................................................GT.AC.T.A.C.TGCC
```



```
N98 rpoBfor.........................................................................................................GT.AC.T.A.C.TGCC
N16 rpoBfor.........................................................................................................GT.AC.T.A.C.TGCC
N62 rpoBfor..................................................................................T......................GT.AC.T.A.C.TGCC
N31 rpoBfor.........................................................................................................GT.AC.T.A.C.TGCC
N64 rpoBfor.........................................................................................................GT.AC.T.A.C.TGCC
N36 rpoBfor.........................................................................................................GT.AC.T.A.C.TGCC
N106 rpoBfo.........................................................................................................GT.AC.T.A.C.TGCC
N37 rpoBfor.........................................................................................................GT.AC.T.A.C.TGCC
N104 rpoBfo.........................................................................................................GT.AC.T.A.C.TGCC
N18 rpoBfor.........................................................................................................GT.AC.T.A.C.TGCC
N111 rpoBfo.........................................................................................................GT.AC.T.A.C.TGCC
N21 rpoBfor.........................................................................................................GT.AC.T.A.C.TGCC
N125 rpoBfo.........................................................................................................GT.AC.T.A.C.TGCC
N44 rpoBfor.........................................................................................................GT.AC.T.A.C.TGCC
N139 rpoBfo.........................................................................................................GT.AC.T.A.C.TGCC
N50 rpoBfor.........................................................................................................GT.AC.T.A.C.TGCC
N113 rpoBfo.........................................................................................................GT.AC.T.A.C.TGCC
N45 rpoBfor.........................................................................................................GT.AC.T.A.C.TGCC
N101 rpoBfo.........................................................................................................GT.AC.T.A.C.TGCC
N46 rpoBfor..................................................................................T......................GT.AC.T.A.C.TGCC
N124 rpoBfo.........................................................................................................GT.AC.T.A.C.TGCC
N42 rpoBfor.........................................................................................................GT.AC.T.A.C.TGCC
N136 rpoBfo.........................................................................................................GT.AC.T.A.C.TGCC
N40 rpoBfor.........................................................................................................GT.AC.T.A.C.TGCC
N110 rpoBfo.........................................................................................................GT.AC.T.A.C.TGCC
N13 rpoBfor.........................................................................................................GT.AC.T.A.C.TGCC
N123 rpoBfo.........................................................................................................GT.AC.T.A.C.TGCC
N25 rpoBfor.................................................................T.......................................GT.AC.T.A.C.TGCC
N133 rpoBfo.........................................................................................................GT.AC.T.A.C.TGCC
N41 rpoBfor.........................................................................................................GT.AC.T.A.C.TGCC
N115 rpoBfo.........................................................................................................GT.AC.T.A.C.TGCC
N20 rpoBfor.........................................................................................................GT.AC.T.A.C.TGCC
N107 rpoBfo.........................................................................................................GT.AC.T.A.C.TGCC
N48 rpoBfor.........................................................................................................GT.AC.T.A.C.TGCC
N47 rpoBfor.........................................................................................................GT.AC.T.A.C.TGCC
N19 rpoBfor.........................................................................................................GT.AC.T.A.C.TGCC
N34 rpoBfor..................................................................................G......................GT.AC.T.A.C.TGCC
N69 rpoBfor.........................................................................................................GT.AC.T.A.C.TGCC
N77 rpoBfor.........................................................................................................GT.AC.T.A.C.TGCC
N82 rpoBfor.........................................................................................................GT.AC.T.A.C.TGCC
N80 rpoBfor.........................................................................................................GT.AC.T.A.C.TGCC
N68 rpoBfor.........................................................................................................GT.AC.T.A.C.TGCC
N92 rpoBfor.........................................................................................................GT.AC.T.A.C.TGCC
N85 rpoBfor.........................................................................................................GT.AC.T.A.C.TGCC
N93 rpoBfor.........................................................................................................GT.AC.T.A.C.TGCC
N73 rpoBfor.........................................................................................................GT.AC.T.A.C.TGCC
N81 rpoBfor.........................................................................................................GT.AC.T.A.C.TGCC
N84 rpoBfor.........................................................................................................GT.AC.T.A.C.TGCC
```


N89 rpoBfor GT.AC.T.A.C.TGCC
N87 rpoBfor GT.AC.T.A.C.TGCC
N58 rpoBfor GT.AC.T.A.C.TGCC
N59 rpoBfor GT.AC.T.A.C.TGCC
N91 rpoBfor GT.AC.T.A.C.TGCC
N70 rpoBfor T GT.AC.T.A.C.TGCC
N74 rpoBfor GT.AC.T.A.C.TGCC
N56 rpoBfor GT.AC.T.A.C.TGCC
N57 rpoBfor GT.AC.T.A.C.TGCC
N63 rpoBfor G GT.AC.T.A.C.TGCC
N83 rpoBfor GT.AC.T.A.C.TGCC
N51 rpoBfor GT.AC.T.A.C.TGCC
N53 rpoBfor GT.AC.T.A.C.TGCC
N66 rpoBfor GT.AC.T.A.C.TGCC
N72 rpoBfor GT.AC.T.A.C.TGCC
N60 rpoBfor GT.AC.T.A.C.TGCC
N71 rpoBfor GT.AC.T.A.C.TGCC
N88 rpoBfor GT.AC.T.A.C.TGCC
N61 rpoBfor GT.AC.T.A.C.TGCC
N75 rpoBfor GT.AC.T.A.C.TGCC
N78 rpoBfor GT.AC.T.A.C.TGCC
N52 rpoBfor GT.AC.T.A.C.TGCC
N54 rpoBfor GT.AC.T.A.C.TGCC
N103 rpoBfc GT.AC.T.A.C.TGCC
N94 rpoBfor GT.AC.T.A.C.TGCC
N126 rpoBfc GT.AC.T.A.C.TGCC
N65 rpoBfor GT.AC.
N128 rpoBfc GT.AC.T.A.C.TGCC
N140 rpoBfc GT.AC.T.A.C.TGCC
N132 rpoBfc GT.AC.T.A.C.TGCC
N137 rpoBfc GT.AC.T.A.C.TGCC
N116 rpoBfc GT.AC.T.A.C.TGCC
N122 rpoBfc GT.AC.T.A.C.TGCC
N127 rpoBfc GT.AC.T.A.C.TGCC
N121 rpoBfc GT.AC.T.A.C.TGCC
N138 rpoBfc GT.AC.T.A.C.TGCC
N100 rpoBfc GT.AC.T.A.C.TGCC
N135 rpoBfc GT.AC.T.A.C.TGCC
N108 rpoBfc GT.AC.T.A.C.TGCC
N117 rpoBfc GT.AC.T.A.C.TGCC
N141 rpoBfc GT.AC.T.A.C.TGCC
N155 rpoBfc GT.AC.T.A.C.TGCC
N163 rpoBfc GT.AC.T.A.C.TGCC
N164 rpoBfc GT.AC.T.A.C.TGCC
N143 rpoBfc GT.AC.T.A.C.TGCC
N156 rpoBfc GT.AC.T.A.C.TGCC
N153 rpoBfc GT.AC.T.A.C.TGCC
N172 rpoBfc GT.AC.T.A.C.TGCC

```
N102 rpoBfc................................................................................................GT.AC.T.A.C.TGCC
N112 rpoBfc................................................................................................GT.AC.T.A.C.TGCC
N114 rpoBfc................................................................................................GT.AC.T.A.C.TGCC
N120 rpoBfc................................................................................................GT.AC.T.A.C.TGCC
N134 rpoBfc................................................................................................GT.AC.T.A.C.TGCC
N149 rpoBfc................................................................................................GT.AC.T.A.C.TGCC
N105 rpoBfc................................................................................................GT.AC.T.A.C.TGCC
N130 rpoBfc................................................................................................GT.AC.T.A.C.TGCC
N129 rpoBfc................................................................................................GT.AC.T.A.C.TGCC
N118 rpoBfc................................................................................................GT.AC.T.A.C.TGCC
N43 rpoBfor................................................................................................GT.AC.T.A.C.TGCC
N131 rpoBfc................................................................................................GT.AC.T.A.C.TGCC
N14rpoBfor.................................................................................................GT.AC.T.A.C.TGCC
N109 rpoBfc................................................................................................GT.AC.T.A.C.TGCC
N119 rpoBfc................................................................................................GT.AC.T.A.C.TGCC
N145 rpoBfc................................................................................................GT.AC.T.A.C.TGCC
N151 rpoBfc................................................................................................GT.AC.T.A.C.TGCC
N152 rpoBfc................................................................................................GT.AC.T.A.C.TGCC
N142 rpoBfc................................................................................................GT.AC.T.A.C.TGCC
N154 rpoBfc................................................................................................GT.AC.T.A.C.TGCC
N144 rpoBfc................................................................................................GT.AC.T.A.C.TGCC
N176 rpoBfc................................................................................................GT.AC.T.A.C.TGCC
N146 rpoBfc................................................................................................GT.AC.T.A.C.TGCC
N158 rpoBfc................................................................................................GT.AC.T.A.C.TGCC
N168 rpoBfc................................................................................................GT.AC.T.A.C.TGCC
N150 rpoBfc................................................................................................GT.AC.T.A.C.TGCC
N162 rpoBfc................................................................................................GT.AC.T.A.C.TGCC
N147 rpoBfc................................................................................................GT.AC.T.A.C.TGCC
N167 rpoBfc................................................................................................GT.AC.T.A.C.TGCC
N157 rpoBfc................................................................................................GT.AC.T.A.C.TGCC
N160 rpoBfc................................................................................................GT.AC.T.A.C.TGCC
N170 rpoBfc................................................................................................GT.AC.T.A.C.TGCC
N161 rpoBfc................................................................................................GT.AC.T.A.C.TGCC
N171 rpoBfc................................................................................................GT.AC.T.A.C.TGCC
N169 rpoBfc................................................................................................GT.AC.T.A.C.TGCC
N166 rpoBfc................................................................................................GT.AC.T.A.C.TGCC
N165 rpoBfc................................................................................................GT.AC.T.A.C.TGCC
N174 rpoBfc................................................................................................GT.AC.T.A.C.TGCC
N159 rpoBfc................................................................................................GT.AC.T.A.C.TGCC
N24 rpoBfor................................................................................................GT.AC.T.A.C.TGCC
N33 rpoBfor................................................................................................GT.AC.T.A.C.TGCC
N35 rpoBfor................................................................................................GT.AC.T.A.C.TGCC
```

Appendix 4: Publications from this PhD work

Appendix 5: IS*6110* FAFLP raw data of 176 Nepal samples _Stored in CD and attached at the back of thesis

Clinical Microbiology and Infectious Diseases

Research Article

ISSN: 2398-5763

Insertion Element IS*6110* based characterisation of Nepalese tuberculosis strains into different genetic lineages

Kartyk Moganeradj[1], **Dunstan Rajendram**[1], **Saraswoti Khadge**[2], **Pam Sonnenberg**[3], **Ibrahim Abubakar**[3,4], **Timothy D. McHugh**[5], **Murdo Macdonald**[2], **Deanna A. Hagge**[2] **and Catherine Arnold**[1]*

[1]Genomic Services and Development Unit, Public Health England, London, UK

[2]Mycobacterial Research Laboratories, Anandaban Hospital, Kathmandu, Nepal

[3]Research Department of Infection & Population Health, University College London, London, UK

[4]Respiratory Diseases Department, TB section, Public Health England, London, UK

[5]Centre for Clinical Microbiology, University College London, London, UK

Abstract

Nepal is geographically located between India and China, a region containing significant Tuberculosis (TB) and Multi-Drug Resistance (MDR-TB) burdens. However, limited information is available on the phylogenetic diversity of Mycobacterium tuberculosis (Mtb) in Nepal. To gain further insight into the diversity of Mtb in Nepal, consecutive clinical samples from 176 newly diagnosed pulmonary tuberculosis patients were collected from two hospitals in Nepal. Insertion Site IS6110 Fluorescent Amplified Fragment Length Polymorphism (FAFLP) PCR and rpoB sequence analysis were carried out on genomic DNA extracts of cultured strains to assign them to accepted genetic lineages and identify MDR-TB. In this study, the IS6110 based characterisation showed a prevalence of 36.36% Central Asian Strain (CAS), 18.75% Beijing, 7.95% Haarlem, 3.97% X, 2.2% each of Latin American Mediterranean (LAM), T-Uganda and T, 1.7% S and 24.4% were unassigned. Further, 3.9% of total M. tuberculosis isolates were of rifampicin resistant genotypes thus indicating that the prevalence of MDR could be higher than the country wide prevalence of MDR among new TB cases (2.2%) as reported by the national drug resistance survey carried out in 2011/2012.

Introduction

TB is ranked as the sixth leading cause of death among the top 20 causes of death in Nepal. According to National Tuberculosis Control Programme (NTCP) in Nepal, in 2014 37,025 TB cases were registered and among them 15,947 (43%) cases were new sputum smear positive TB cases. It was estimated by WHO [1] that 4.6 (2.1-7.5) thousand people in Nepal died from TB in 2014. Even though short course TB drug treatment regimen could cure around 89% of cases, TB mortality was still unacceptably high in Nepal. Since 2006, the STOP TB strategy has been adopted by NTCP. However Drug resistant TB (DR-TB) still threatens national TB control and is a major public health concern. The proportion of MDR-TB cases in new cases was 2.2% and retreatment cases were 15.4%. Even though the Millennium Development Goal (MDG) to halt and reverse TB incidence has been achieved in all six WHO regions, work remains to be done to prevent the deaths from this dreadful disease [1].

The identification of the number and position of Insertion Sequence IS*6110* elements in the Mtb genome has been widely used as a genomic tool for the rapid fingerprinting of isolates of *Mycobacterium tuberculosis* complex (MTBC) [2]. IS*6110* based Restriction Fragment Length Polymorphism (RFLP) is considered as the 'gold standard' typing method for strains with more than five copies [3-5]. As IS*6110* transposition is among the first genetic changes to occur in strains from a transmission chain [6], this marker has also been used for outbreak analysis [5].

Modification of the conventional IS*6110* typing method, using differentially labelled primers has allowed characterisation of Mtb isolates into the key genetic lineages more rapidly than traditional methods [7]. This approach can be facilitated with automation, which enables this technique to be performed in a high throughput setting. The fragment patterns generated indicates both copy number and insertion site of IS*6110* in the genome [8,9]. The patterns generated correlates directly with other independent markers and can be used for transmission investigation locally or internationally. Specific fragments are common in genetically related lineages and do not occur in other groups (e.g. spoligotype groups such as Beijing and the Euro-American lineage which contain the Latin American Mediterranean (LAM), Haarlem, S, T and X spoligotype groups). The patterns generated correlate directly with other independent markers and can be used for transmission investigation locally/internationally. Principal Genetic Groups (PGGs) can be assigned to Mtb strains based on the combination of polymorphism located at *katG* codon463 and *gyrA* codon95 in the respective genomes [10] or spoligotypes [11] or global phylogeny classification based on whole genome sequences [12].

Correspondence to: Catherine Arnold, Genomic Services and Development Unit, Public Health England, London, UK, Tel: 0208-327-6068; **E-mail:** Catherine.arnold@phe.gov.uk

Key words: *tuberculosis, IS6110, FAFLP, rifampicin resistance, TB lineages*

Received: June 01, 2016; **Accepted:** June 27, 2016; **Published:** June 30, 2016

Limited data are available on the characterisation of Mtb strains and genotypes circulating in Nepal. A key factor is the geographical location of Nepal, interlocked between China and India, two countries, which together account for approximately a third of annual global new cases (11% and 24%, respectively) [1]. A recent study of 261 Nepalese isolates found any drug resistance (any drug resistance has been defined as resistance to isoniazid, rifampicin, streptomycin, ethambutol, fluoroquinolones, and/or aminoglycosides) in 12.8%% of Mtb strains that were new untreated cases, with the most frequent lineages reported as CAS/Delhi (40.6%), East Asian (including Beijing) (32.2%), Euro-American (15.7%) and Indo-oceanic (11.5%) [13]. To gain further insight into the characteristics and diversity of Mtb in Nepal, our study aimed to categorise isolates for the first time using IS*6110* FAFLP PCR and to assign them to different genetic lineages. Secondly, the level of MDR would be characterized in the population using *rpo*B (*rpo*B gene encodes for the Beta subunit of bacterial RNA polymerase) sequencing of the Rifampicin Resistant Determining Region (RRDR) as a predictive surrogate [14-17].

Methods

Strains

Sputum samples from 176 consecutive new TB patients over one year were collected between 2007 and 2008 and cultured alongside routine diagnostics from two Nepalese tuberculosis reference centres located in the Kathmandu valley: the National Tuberculosis Centre (NTC) and the German Nepal Tuberculosis Project (GENETUP). The patient population represented local and referred cases from across Nepal. Bacterial genomic DNA from isolated strains was extracted by the Cetyltrimethylammonium Bromide (CTAB) method [18] at the Mycobacterial Research Laboratories (MRL) in Anandaban Hospital. Informed consent was not required at the time of this study, as samples were collected with routine clinical care and all patient identifiers were anonymized; however, all patients were provided an explanation and were only included upon provision of verbal informed consent. Study procedures were reviewed and approved by NTC and GENETUP. The results for the drug sensitivity tests were unavailable during the entire duration of this study.

IS*6110* FAFLP PCR, Fragment Sizing and Analysis

Genomic DNA was digested with the restriction enzymes *MseI* and *TaqI* followed by ligation with double stranded *TaqI* restriction site specific adaptors. The adaptor ligated DNA was amplified following previously published PCR conditions using four fluorescently labelled adaptor specific *TaqI* forward primers - 5'-CGATGAGTCCTGACCGA*/C*/T*/G* each labelled with a single unique selective nucleotide at the 3' end and an IS*6110* sequence specific reverse primer- 5'- CTGACATGACCCCATCCTTT [9]. In a total volume of 20 µl, 1 µl of the adaptor ligated DNA was added to the reaction containing 1X reaction buffer, 1.5 mM MgCl$_2$, 0.2 mM dNTPs (Invitrogen, UK) , 1 µM of labelled *Taq I* forward primer, 1 µM of IS*6110* reverse primer and 1 U of recombinant *Taq* polymerase (Invitrogen, UK). The following PCR conditions were carried out in a Veriti thermocycler (Applied Biosystems, UK): 94°C for 15min followed by 35 cycles of 94°C for 20s, 66°C for 30s and 72°C for 2 min with the 66°C annealing temperature reducing by 1°C every cycle for nine cycles and the last 25 cycles at 56°C. Finally, an extension of 72°C for 60min was carried out before further manipulations. The fragments were separated on an ABI genetic analyser 3730XL (Applied Biosystems, UK), sized using PeakScanner v1.0 software (Applied Biosystems) and identified using their fluorescent tag (Figure 1). The four-dye FAFLP data collected from the different profiles were then recorded and compared with a reference collection of Mtb isolates [19] using BioNumerics software v6.1 (Applied Maths Inc., Belgium). Fragments common to different lineages (defined as being present in >50% of strains in a particular genetic lineage) were recorded for each Nepalese strain and compared with a fully characterised global collection as detailed by Thorne *et al.* [20]. These data were then used to build a dendrogram using the Dice coefficient of similarities to compare the similarity matrix and Unweighted Pair Group Method with Arithmetic Mean (UPGMA) derived cluster analysis with cophenetic correlation for the branch quality.

*rpo*B analysis

The 81 bp Rifampicin Resistant Determining Region (RRDR) of the *rpo*B gene of all strains were sequenced using published primers [21] and analysed in BIOEDIT software using ClustalW alignment parameters. The PCR was carried out in a total volume of 50 µl where 1 µl of the DNA was added to the reaction containing 1xPCR reaction buffer, 1.5 mM MgCl$_{2,}$ 0.2 mM dNTPs (Invitrogen, UK), 20 µM each of both *rpo*B-RRDRforward (5'-CGATCACACCGCAGACGTTGA) and reverse primers (5'- GGCACGCTCACGTGACAGACC) and 5U recombinant Taq polymerase (Invitrogen, UK). The following PCR conditions were carried out using a Veriti thermocycler (Applied Biosystems, UK): 94°C for 2 min followed by 35 cycles of 94°C for 30 sec, 60°C for 30 sec and 72°C for 1 min. Finally, an extension of 72°C for 10 min was performed before cleaning the products using AmpureXP magnetic beads (Beckman Coulter, UK) according to the manufacturer's protocol and sequencing using the forward primer, *rpo*B-RRDR forward.

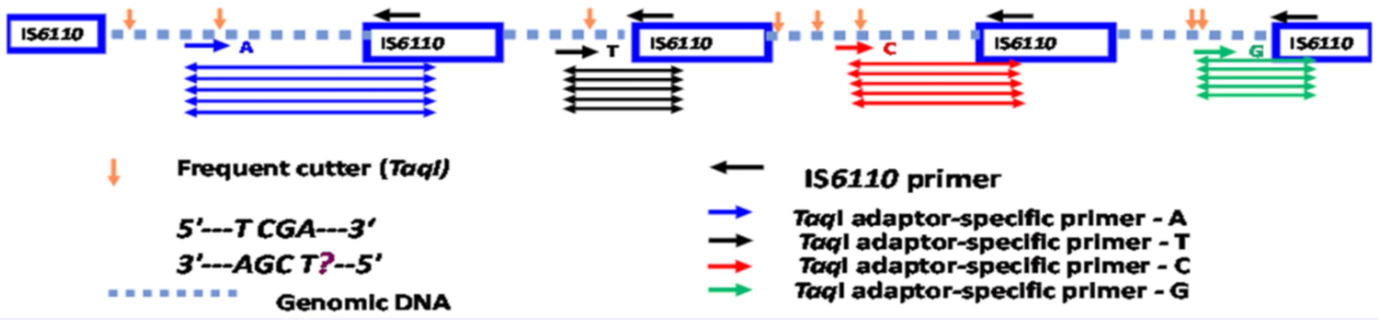


Figure 1. Pictorial representation of the IS*6110* FAFLP PCR methodology. Coloured fragments are amplified using uniquely labelled adapter specific primers. For example, blue coloured fragments are amplified if the primers are labelled with 6-FAM fluorescein 'blue' dye, red coloured fragments with PET dye, green coloured fragments with VIC dye and black coloured fragments with NED dye.

Results

Analysis of Data using BioNumerics software v6.1

Of the 176 DNA extracts from isolates analysed, the majority of the samples 97 (55.4%) belonged to either the spoligotype-defined Central Asian Strain (CAS) lineage (64 *i.e.,* 36.6%) or the Beijing lineage (33 *i.e.,* 18.8%) grouping under PGG1 and the rest of the samples group under either PGG2 (1.7% S, 3.97% X, 7.95% Haarlem and 2.27% LAM, 2.27% T-Uganda) or PGG3 (2.27% of T) (Table 1). Forty three samples (24.4%) grouped under "unassigned" group. Common fragments seen were exactly the same as the earlier published report by Thorne *et al.,* (2011) except for an additional fragment, 78.4 G, for the CAS lineage. A dendrogram was generated using only the IS*6110* FAFLP data (Figure 2) confirming again the above mentioned lineages in relation to the PGGs.

*rpo*B *Analysis*

Of 176 DNA extracts analysed for *rpo*B mutations, seven samples (3.9%) had a single non-synonymous base change which would likely confer resistance to rifampicin (Table 2). Six of these seven samples showed a second base mutation in a codon triplet whereas sample N70 showed a first base mutation.

Discussion

It has been demonstrated previously that IS*6110* FAFLP PCR can be used to delineate the phylogeny of MTBC as shared common fragments can determine the different lineages in a geographical location by comparison with a reference database collection [7]. As limited lineage information is available from strains in Nepal, we have applied the IS*6110* method published recently on mapping the IS*6110* sites in H37Rv [9] and also carried out rpoB sequencing to further characterize strains from this important region.

Fifty five percent of the 176 Nepalese strains analysed belong to the CAS (36.6%) and Beijing (18.8%) modern genetic spoligotypes (PGG1). The remaining 24.4% of the samples belong to the PGG2 and PGG3 groups (Haarlem, LAM, S, X, T-Uganda and T). However, a limitation of this technique is its difficulty to characterise the samples with less than 4-5 copies of IS*6110* as seen in the unassigned group (24.4%) in figure 2, which can be overcome by the use of other typing techniques like Mycobacterial Interspersed Repetitive Units- Variable Number Tandem Repeats (MIRU-VNTR) [20]. The geographical position of Nepal is likely to have influenced this distribution, with a mixture of predominantly Beijing lineage from the North of the Himalayas and the CAS lineage from the south [22]. The fairly high

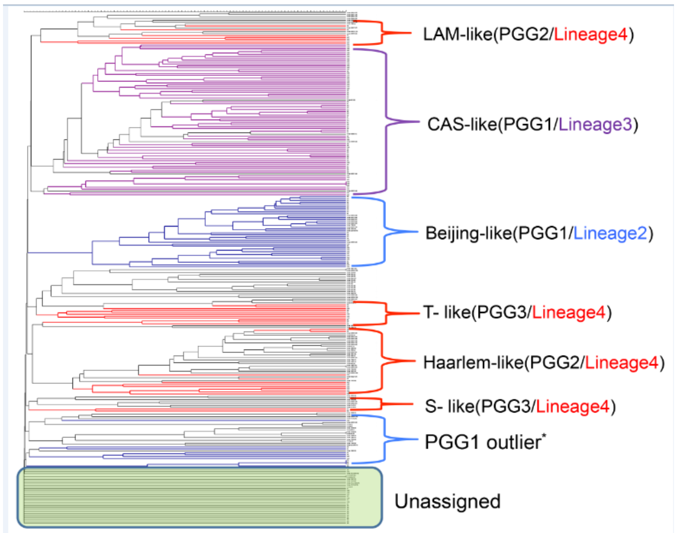


Figure 2. UPGMA derived dendrogram showing the predominant genetic lineages / spoligotypes of 176 Nepalese Mtb isolates.
Coloured branches represent Nepal samples following Gagneux's global phylogeography of MTBC [12] and black branches represent the in-house global Mtb collection. PGG1 outliers share one IS*6110* copy with the PGG1 group. Those which contained one IS*6110* copy but could not be assigned to any group (unassigned group in the figure) are shown within the green box.

Table 1: Common fragments identified using IS*6110* FAFLP PCR in TB genetic lineages between the 176 bacterial DNA isolates in Nepal.

PGG /spoligotype/ sub-lineage	Common fragment sizes	No. Of Nepal strains
PGG1-CAS	78.4G ,92.0B , 117.9R, 206.2G , 275.1R	64/176
PGG1-Beijing	101.7B, 102.5Y, 139.1R, 180.7Y, 254.8G, 332.4R, 353.5B	33/176
PGG2 –Haarlem	87.0Y, 89.4G, 148.7B, 300.2R (H/X), 445.7Y	14/ 176
PGG2-LAM	71.5Y, 105.2R, 116.1R	4/ 176
PGG2-S	88.0G, 112.9R, 217.2R, 445.3G	3/ 176
PGG2-X	83.8Y, 300.2R(H/X)	7/ 176
PGG2-T Uganda	88.9Y, 119.5G, 122.9Y, 228.4Y, 266.8R	4/ 176
PGG3- T	81.3R, 192.4R, 360R	4/176
Ungrouped		43/176

Where B-Blue coloured fragment R-Red coloured fragment G-Green coloured fragment and Y-Black/ Yellow coloured fragment seen in the electropherogram. PGG represents Principal Genetic groups according to Sreevatsan et al. [10], spoligotypes follow spolDB4 classification[4] and sub-lineages are grouped following Gagneux's classification [8].

Table 2: List of mutations seen in *rpo*B Rifampicin Resistance-Determining Region (RRDR) of rifampicin resistant *M. tuberculosis* isolates from Nepal.

Sample	Mutated locus	Nucleotide modification	Amino acid modification	FAFLP derived
1. N70	516	GAC > TAC	(Asp>Tyr)	PGG1- Beijing group
2. N10	522	TCG > TTG	(Ser>Leu)	PGG1- CAS group
3. N25	526	CAC > CTC	(His>Leu)	PGG3 -like
4. N63	526	CAC > CGC	(His>Arg)	PGG3-like
5. N34	531	TCG > TGG	(Ser>Trp)	PGG3- T group (H37Rv-like)
6. N46	531	TCG > TTG	(Ser>Leu)	PGG3 (H37Rv-like)
7. N62	531	TCG > TTG	(Ser>Leu)	Ungrouped

* PGG represents Principal Genetic Group.

percentage of mainly European lineages (Haarlem, LAM and T, 12.5%) indicates that there has also been mixing of the different lineages over an extended time and that European travellers/migrants to South East Asia and Nepal may have transmitted European strains to the local population. According to Malla *et al.* [13], fifty strains had any drug resistance and sixteen (6.1%) out of 261 isolates were MDR. Among the fifty any drug resistant strains, 29 cases were previously treated and twenty-one were new untreated cases (8.04% of 261 total strains and 12.8% of 164 new untreated cases). In this study, MDR-TB was tested by using rifampicin as the resistance marker and 7 isolates out of 176 were found (3.9%) from new untreated cases possessing drug resistance genotypes. However, their reported MDR percentage was based on the total number of isolates, of which 37.2% of isolates were from previously treated cases, which may enhance a probability of drug resistance development compared to untreated cases. Our results indicate that the prevalence of Rifampicin (RIF) resistant TB (surrogate marker for MDR) was higher than the nationally reported 2.2% MDR in new untreated cases. Further, our results are concordant with a recently conducted study by Creswell *et al.* [23], where they have shown that the genotypic rifampicin resistance in newly diagnosed TB patients to be 3.3% in Nepal [23].

Conclusions

The IS*6110* FAFLP data from our study reiterates the fact that the geographic location of Nepal is the key for the circulation of PGG1 TB lineages, CAS and Beijing, which were predominant in India and China respectively. Further the RRDR study correlates with the recent work by Creswell *et al.* showing that prevalence of MDR-TB may be marginally higher than the national average in new untreated TB cases. As the monitoring of TB is important in Nepal, this simple and informative PCR-based molecular epidemiological technique would prove useful for the study of outbreaks of the disease and also to detect cross-contamination between different strains or isolates in resource poor settings. The most common mutation site in the RRDR is at codon 531 and parallels the findings of earlier studies [21,22].

Authorship

All authors mentioned above gave substantial intellectual contribution to this manuscript.

Acknowledgments

We thank gratefully to the staff of the NTC and GENETUP from the Kathmandu valley in Nepal. We would like to thank all staffs from MRL, Anandaban Hospital in Lalitpur, Nepal for their technical help. We are also grateful to Dr. Ali Al-Shahib from Bioinformatics Department (PHE, UK) and Mr. Jon White from Medical Illustration Department (PHE, UK) for their technical expertise.

Funding information

The authors received no specific funding for this work.

Competing interest

The authors declare that they have no competing interests.

References

1. WHO (2014) Global tuberculosis report 2014.
2. Haas WH, Butler WR, Woodley CL, Crawford JT (1993) Mixed-Linker Polymerase Chain Reaction?: a New Method for Rapid Fingerprinting of Isolates of the Mycobacterium tuberculosis Complex. *Tuberculosis* 31: 1293-1298. [Crossref]
3. Otal I, Martín C, Vincent-Lévy-Frebault V, Thierry D, Gicquel B (1991) Restriction fragment length polymorphism analysis using IS6110 as an epidemiological marker in tuberculosis. *J Clin Microbiol* 29: 1252-1254. [Crossref]
4. de Boer AS, Borgdorff MW, de Haas PE, Nagelkerke NJ, van Embden JD, et al. (1999) Analysis of rate of change of IS6110 RFLP patterns of Mycobacterium tuberculosis based on serial patient isolates. *J Infect Dis* 180: 1238-1244. [Crossref]
5. van Embden JD, Cave MD, Crawford JT, Dale JW, Eisenach KD, et al. (1993) Strain Identification of Mycobacterium tuberculosis by DNA Fingerprinting?: Recommendations for a Standardized Methodology. *J Clin Microbiol* 31: 406-409. [Crossref]
6. Schürch AC, Kremer K, Kiers A, Daviena O, Boeree MJ, et al. (2010) The tempo and mode of molecular evolution of Mycobacterium tuberculosis at patient-to-patient scale. Infection, genetics and evolution? *Infect Genet Evol* 10: 108-114. [Crossref]
7. Thorne N, Borrell S, Evans J, Magee J, García de Viedma D, et al. (2011) IS6110-based global phylogeny of Mycobacterium tuberculosis. Infection, genetics and evolution. *Infect Genet Evol* 11: 132-138. [Crossref]
8. Borrell S (2009) Comparison of four-colour IS6110-fAFLP with the classic IS6110-RFLP on the ability to detect recent transmission in the city of Barcelona, Spain. *Tuberculosis* (Edinburgh, Scotland) 89: 233-237. [Crossref]
9. Moganeradj K, Abubakar I, McHugh TD, Sonnenberg P, Arnold C (2013) Insertion site mapping for repeated elements in Mycobacterium tuberculosis. *J Microbiol Methods* 92: 192-196. [Crossref]
10. Sreevatsan S, Pan X, Stockbauer KE, Connell ND, Kreiswirth BN, et al. (1997) Restricted structural gene polymorphism in the Mycobacterium tuberculosis complex indicates evolutionarily recent global dissemination. *Proc Natl Acad Sci USA* 94: 9869-9874. [Crossref]
11. Brudey K, Driscoll JR, Rigouts L, Prodinger WM, Gori A, et al. (2006) Mycobacterium tuberculosis complex genetic diversity?: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. *BMC Microbiol* 6: 23. [Crossref]
12. Gagneux S (2012) Host-pathogen coevolution in human tuberculosis. *Philos Trans R Soc Lond B Biol SciPhilosophical* 367: 850-859. [Crossref]
13. Malla B, Stucki D, Borrell S, Feldmann J, Maharjan B, et al. (2012) First insights into the phylogenetic diversity of Mycobacterium tuberculosis in Nepal. *PLoS One* 7: e52297. [Crossref]
14. Luita Nice CO, Jairo da Silva MS, Luiz Alexandre VM, Sônia Cristina OM, Antonio M (2016) Detection of multidrug-resistant Mycobacterium tuberculosis strains isolated in Brazil using a multimarker genetic assay for katG and rpoB genes. *Braz J Inf Dis* 20: 166-172. [Crossref]
15. Tripathi R, Sinha P, Kumari R, Chaubey P, Pandey A, et al. (2016) Detection of rifampicin resistance in tuberculosis by molecular methods: A report from Eastern Uttar Pradesh, India. *Indian J Med Microbiol* 34: 92-94. [Crossref]
16. Suresh N, Singh UB, Arora J, Pande JN, Seth PSuresh, N, et al. (2006) Rapid detection of rifampicin-resistant Mycobacterium tuberculosis by in-house, reverse line blot assay. *Diag Microbiol Infec Disease* 56: 133-140. [Crossref]
17. Somoskovi A, Parsons LM, Salfinger M (2001) The molecular basis of resistance to isoniazid, rifampin, and pyrazinamide in Mycobacterium tuberculosis. *Respir Res* 2: 164-168. [Crossref]
18. van Soolingen D, Hermans PW, de Haas PE, Soll DR, van Embden JD (1991) Occurrence and stability of insertion sequences in Mycobacterium tuberculosis complex strains: evaluation of an insertion sequence-dependent DNA polymorphism as a tool in the epidemiology of tuberculosis. *J Clin Microbiol* 29: 2578-2586. [Crossref]
19. Kristin K, Catherine A, Angel C, Cristina MG, Walter HH (2005) Discriminatory Power and Reproducibility of Novel DNA Typing Methods for Mycobacterium tuberculosis Complex Strains Discriminatory Power and Reproducibility of Novel DNA Typing Methods for Mycobacterium tuberculosis Complex Strains. *J Clin Microbiol* 43: 5628-5638. [Crossref]
20. Thorne N, Evans JT, Smith EG, Hawkey PM, Gharbia S, et al. (2007) An IS6110-targeting fluorescent amplified fragment length polymorphism alternative to IS6110 restriction fragment length polymorphism analysis for Mycobacterium tuberculosis DNA fingerprinting. *Clin Microbiol Infect* 13: 964-970. [Crossref]
21. Arnold C, Westland L, Mowat G, Underwood A, Magee J, et al. (2005) Single-nucleotide polymorphism-based differentiation and drug resistance detection in

Mycobacterium tuberculosis from isolates or directly from sputum. *Clin Microbiol Infect* 11: 122-130. [Crossref]

22. Ajay P, Chie N, Yukari F, Haruka S, Basu DP, et al. (2012) Molecular characterization of multidrug-resistant Mycobacterium tuberculosis isolated in Nepal. *Antimicrob Agents Chemother* 56: 2831-2836. [Crossref]
23. Creswell J, Rai B, Wali R, Sudrungrot S, Adhikari LM, et al. (2015) Introducing new tuberculosis diagnostics: the impact of Xpert(®) MTB/RIF testing on case notifications in Nepal. *Int J Tuberc Lung Dis* 19: 545-551. [Crossref]

PLOS | ONE

RESEARCH ARTICLE

Sequence analysis of the rifampicin resistance determining region (RRDR) of rpoB gene in multidrug resistance confirmed and newly diagnosed tuberculosis patients of Punjab, Pakistan

Salma Hameed[1,2], Kartyk Moganeradj[3]*, Nasir Mahmood[4], Timothy D. McHugh[2], Muhammad Nawaz Chaudhry[1], Catherine Arnold[3]

1 College of Earth and Environmental Sciences, University of the Punjab, Lahore, Pakistan, **2** Centre for Clinical Microbiology, Division of Infection and Immunity, Royal Free Campus, University College London, London, United Kingdom, **3** Genomic Services and Development Unit, Public Health England, Colindale, United Kingdom, **4** Department of Biochemistry; Department of Human Genetics and Molecular Biology, University of Health Sciences, Lahore, Pakistan

* Kartyk.moganeradj@phe.gov.uk

Citation: Hameed S, Moganeradj K, Mahmood N, McHugh TD, Chaudhry MN, Arnold C (2017) Sequence analysis of the rifampicin resistance determining region (RRDR) of rpoB gene in multidrug resistance confirmed and newly diagnosed tuberculosis patients of Punjab, Pakistan. PLoS ONE 12(8): e0183363. https://doi.org/10.1371/journal.pone.0183363

Editor: Dipankar Chatterji, Indian Institute of Science, INDIA

Received: May 17, 2017

Accepted: July 24, 2017

Published: August 17, 2017

Data Availability Statement: All relevant data are within the paper and its Supporting Information file.

Funding: This work was supported by Higher Education Commission, Pakistan (grant no: 1-8/HEC/HRD/2015/3697 PIN: IRSIP 28 BMS13).

Competing interests: The authors have declared that no competing interests exist.

Abstract

Molecular screening of new patients suspected for TB could help in the effective control of TB in Pakistan as it is a high TB burden country. It will be informative to understand the prevalence of multi drug resistance for a better drug regimen management in this geographical area. The Rifampicin resistance determining region (RRDR) sequencing was used to identify mutations associated with drug resistance in DNA extracts from 130 known multidrug resistant (MDR) cultured strains and compared with mutations observed in DNA extracts directly from 86 sputum samples from consecutive newly diagnosed cases in Lahore, Pakistan. These newly diagnosed samples were positive for smear microscopy, chest X-ray and presumed sensitive to first line drugs. In the known MDR group the most frequent mutations conferring resistance were found in *rpo*B531 (n = 51, 39.2%). In the newly diagnosed tuberculosis group with no history of MDR, mutations in *rpo*B531 were seen in 10 of the samples (11.6%). Collectively, all mutations in the RRDR region studied were observed in 80 (61.5%) of known MDR cases and in 14 (16.3%) of the newly diagnosed cases. Using the RRDR as a surrogate marker for MDR, sequences for the newly diagnosed (presumed sensitive) group indicate much higher levels of MDR than the 3.9% WHO 2015 global estimate and suggests that molecular screening directly from sputum is urgently required to effectively address the detection and treatment gaps to combat MDR in this high burden country.

Introduction

Pakistan is among the top 20 countries with a high TB and MDR-TB burden [1]. Chest X-ray, acid fast bacilli (AFB) smear microscopy and culture on Lowenstein-Jensen (LJ) media are the conventional methods of investigation for tuberculosis [2] but require additional analysis to define the species of Mycobacteria as well as the mechanism of drug resistance. *M. tuberculosis* drug resistance detection using conventional methods is by culture of bacilli on a medium containing antibiotic and can require several weeks. However, with the development of rapid molecular methods it is possible to detect mutations in genes associated with resistance in a much shorter time [3]. Resistance to two first line predominant anti-TB drugs i.e., isoniazid (INH) and rifampicin (RIF), is termed as 'multidrug resistance tuberculosis' [4]. Molecular methods for this are diverse and each method has its benefits and drawbacks; for example PCR-RFLP [5] and allele-specific PCR [6]. Several molecular techniques have been evolved to detect the gene mutation related to resistance. These include hybridization methods; single strand polymorphism, DNA sequencing and other PCR based methods [5, 7, 8]. Multiplex Allele Specific (MAS) PCR, a rapid and cost-effective method simultaneously detects INH, RIF and Ethambutol (EMB) resistance associated genetic mutations [9]. PCR technology can provide many advantages over traditional techniques. Many PCR tests can be rapidly performed and interpreted on the same day of submission of samples. A major advantage of PCR over traditional techniques includes the ability to rapidly identify organisms that are difficult to culture and the DNA of interest can be amplified with the DNA from just one cell. The sensitivity of PCR is also its major disadvantage since very small amounts of contaminating DNA (from a different sample) can also be amplified. One major limitation of PCR is that prior information about the target sequence is necessary in order to generate the primers that will allow its selective amplification [10]. Whole Genome Sequencing (WGS) sequences the whole genome rather than specific genes. So, drug resistance prediction from the whole genome sequence is possible using publically available software which rapidly analyses all known gene targets and identifies mutations associated with resistance thus enabling targeted treatment [11] but requires culture and is currently prohibitively expensive for high burden countries. Identification of MDR-TB is a crucial step as treatment of multi drug resistant tuberculosis (MDR-TB) is a considerable challenge. Globally in 2015, an estimated 3.9% (95% confidence interval [CI]: 2.7–5.1%) of new cases and 21% (95% CI: 15–28%) of previously treated cases had MDR/RR-TB [1]. Resistance to rifampicin is the result of mutations in the rifampicin resistance determining region (RRDR) of *rpo*B, particularly mutations at codons 516, 526 and 531. MDR-TB is defined as resistance to rifampicin and isoniazid, the two most effective anti-TB drugs. In December 2010, WHO recommended the use of the GeneXpert MTB/RIF to detect and infer resistance to rifampicin directly from sputum [12]. In May 2016, WHO issued guidance that "people with TB resistant to rifampicin, with or without resistance to other drugs, should be treated with an MDR-TB treatment regimen." Together with MDR-TB, these are referred to as MDR/RR-TB.

This work was carried out in Pakistan to characterize mutations associated with rifampicin resistance directly from sputum samples from newly diagnosed (ND) patients with no history of drug resistance and identify their key risk factors in this setting. Identifying resistance in presumed resistant samples and inferring resistance profiles directly from sputum may enable a better tailored drug regimen where possible. It will also inform patient management more rapidly and consequently reduce the rate of onward transmission of MDR tuberculosis in high burden countries such as Pakistan.

Materials and methods

Sampling

Tuberculosis patients attending the Ghulab Devi Chest Hospital, Lahore, Pakistan in collaboration with University of Health Sciences, Lahore were enrolled in this study over 18 months between 2013 and 2015, based on the following inclusion criteria. For the MDR group, patients diagnosed previously with TB, and with a history of resistance to first line anti-tuberculosis drugs, were included. Sputum samples were taken from this group and culture and Drug Susceptibility Testing (DST) were performed. The second group of patients included freshly diagnosed consecutive cases, presumed drug susceptible with clinical symptoms of TB, positive in sputum smear microscopy AFB, chest X-ray positive and no history of resistance to any first line tuberculosis drugs. Sputum samples were taken from this group and culture was not performed. Not all individuals were included; those with clinical complications in addition to tuberculosis were not included in this study.

Patient history

The patient's history was collected using a proforma and included age, gender, area, economic status (earning less than 300 US dollars per month), information of previous anti-tuberculosis therapy, chest x-ray, AFB test and family history of TB. The environmental parameters studied were animal contact, source of drinking water, un-boiled milk use and smoking or drug use.

GeneXpert testing

GeneXpert testing was carried out only for six samples of the MDR group only due to lack of global funding, according to the manufacturer's instructions. Newly diagnosed presumed susceptible samples were not tested due to the reason that these samples have no history of drug resistance. This test was performed in order to confirm the samples of MDR group before carrying out DST.

Initial sputum culturing on drug free LJ medium

For the MDR group, sputum suspension of each patient was made by mixing 0.5ml sputum in equal volume of autoclaved deionized water under aseptic conditions. 0.1ml of sputum sample was spread on LJ medium for *M. tuberculosis* culture under strict aseptic conditions. The colonies appeared on the LJ medium after 4–6 weeks of culturing at 37˚C.

Drug susceptibility testing (DST)

Sub culturing of *M. tuberculosis* colonies from the MDR group was carried out on LJ medium containing the different first line anti-tuberculosis drugs. The concentrations of drugs added were: rifampicin (40 μg/ml); isoniazid (0.2 μg/ml); ethambutol (2 μg/ml); pyrazinamide (50 μg/ml) andstreptomycin (4 μg/ml). Following a sterility check by incubating the culture bottles at 37˚C for one week, the bottles were inoculated with an *M. tuberculosis* suspension of the previous culture. After 4–6 weeks incubation at 37˚C in incubator, growth on a drug-free control medium was compared with growth on culture media containing each concentration of anti-tuberculosis drug. Any growth of *M. tuberculosis* colonies on drug containing LJ media were designated drug resistant while samples where no growth was observed on drug containing LJ media, were declared to be drug susceptible *Mycobacterium tuberculosis*.

DNA isolation and quantification

Sputum samples from the presumed susceptible group of patients were collected and then DNA extraction was carried out. Both cultures (described above) and sputum samples were extracted using the column based TIANamp genomic DNA isolation kit (TIANGEN Biotech Beijing, China) method. Quantity and quality of the isolated genomic DNA was determined by NanoDrop (Thermo Scientific, USA) using 1μL sample of purified DNA.

*rpo*B analysis

The fragment containing 81bp Rifampicin Resistant Determining Region (RRDR) of the *rpo*B gene of all strains were sequenced using published primers [13] and analysed in BIOEDIT software using ClustalW alignment parameters (BioEdit version 7.2.5). The PCR was carried out in a total volume of 50μl where 1μl of the DNA was added to the reaction containing 1xPCR reaction buffer, 1.5mM $MgCl_2$, 0.2mM dNTPs (Invitrogen, UK), 20μM each of both *rpo*B-RRDR forward (5′ – CGATCACACCGCAGACGTTGA) and reverse primers (5′ –GGCAC GCTCACGTGACAGACC) and 5U recombinant Taq polymerase (Invitrogen, UK). The following PCR conditions were carried out using a Veriti thermocycler (Applied Biosystems, UK): 94˚C for 2 min followed by 35 cycles of 94˚C for 30 sec, 60˚C for 30 sec and 72˚C for 1 min. Finally, an extension of 72˚C for 10 min was performed before cleaning the products using AmpureXP magnetic beads (Beckman Coulter, UK) according to the manufacturer's protocol and forward and reverse sequencing performed.

Ethics and consent

The present research work was approved by the ethical committee of University of the Punjab, Lahore, Pakistan in accordance with the ethical standards of the responsible committee on human experimentation and with the latest (2008) version of Helsinki Declaration of 1975 [14]. The purpose of the study was explained and written consents from the patients or guardians were taken from all patients or from next of their kin, caretakers, or guardians/parents on behalf of all child participants.

Results

GeneXpert testing

GeneXpert testing was positive for six of the samples of MDR group and further verified by DST.

Drug susceptibility testing (DST)

Of the 130 MDR cultures tested for resistance to isoniazid (I), rifampicin (R), ethambutol (E), pyrazinamide (P) and streptomycin (S), 96 were resistant to IREPS, 26 were resistant to IREP, three were resistant to IR, two were resistant to IRES, one was resistant to IRP and two resistant to IRPS. Streptomycin was added so that to kill any other bacterial contamination in the culture in addition to drug sensitive mycobacteria.

*rpo*B analysis

An overview of the *rpo*B mutations seen in both groups is shown in Table 1. The isolated DNA from all samples quantification was carried out by nanodrop and DNA quantity was found to

Table 1. Mutations seen in rifampicin resistance determining region (RRDR).

	516 (WT = GAC- Asp)			526 (WT = CAC- His)			531 (WT = TCG- Ser)		
SNP	***T*AC**	**G*T*C**	**G*G*C**	***T*AC**	***A*AC**	**C*C*C**	**T*T*G**	**T*G*G**	**T*G*C**
(AA change)	**(Tyr)**	**(Val)**	**(Gly)**	**(Tyr)**	**(Asn)**	**(Pro)**	**(Leu)**	**(Trp)**	**(Cys)**
MDR group (n = 130)	8	7	2	4	2	1	49	1	1
ND group (n = 86)	2	2	0	0	0	0	8	2	0

Note: Multidrug resistant group (MDR), Newly diagnosed group (ND).

https://doi.org/10.1371/journal.pone.0183363.t001 be in range of 70–90 ng/μl. Of the 130 MDR strains, 80 had mutations in the RRDR region of rpoB (61.3%) (S1 Fig). In order of mutation frequency, 49 strains carried a single mutation at position 531, TCG>TTG/Ser>Leu (65.3%), 8 strains carried a single mutation at position 516, GAC>TAC/Asp>Tyr (10%), 7 strains carried a single mutation at position 516 GAC>GTC/Asp>Val (8.8%), 4 strains carried a single mutation at position 526, CAC>TAC/His>Tyr (5%), 2 strains carried a single mutation at position 516, GAC>GGC/Asp>Gly (2.5%), a further 2 strains carried a single mutation at position 526, CAC>AAC/His>Gly (2.5%) and single strains carried mutations at positions 526, CAC>CCC/His>Pro; 531, TCG>TGG/Ser>Trp; TCG>TGC/Ser>Cys (all at 1.3%). The remaining mutations were as follows: one strain with a deletion of positions 516 and 517; one strain with two mutations, the first at position 516, GAC>GGC/Asp>Ala and the second at position 531, TCG>GCG/Ser>Ala; one strain with two mutations, the first at position 526, CAC>CAG/His>Gln and the second at position 533, CTG>CCG/Leu>Pro; one strain with a two amino acid deletion of 516/7 and two strains with mutations just upstream of the RRDR.

For the ND group, 14/86 of the extracted DNAs (16.3%) carried mutations associated with MDR status; the remainder showed wild type RRDR sequence. Of the 14, 8 DNAs carried a single mutation at position 531, TCG>TTG/Ser>Leu (57.1%); 2 DNAs carried a single mutation at position 531, TCG>TGG/Ser>Trp (14.3%); 2 DNAs carried a single mutation at position 516, GAC>GTC/Asp>Val (14.3%) and a further 2 DNAs carried a single mutation at position 516, GAC>TAC/Asp>Tyr (14.3%).

Discussion

Molecular screening of *M. tuberculosis*–containing sputum samples for drug resistance, although recommended by the WHO, is expensive and inaccessible to many high incidence areas such as Pakistan. To gain further information about the prevalence of MDR in newly diagnosed patients in this area of Lahore, Pakistan, the RRDR region of the *rpo*B gene from two groups was sequenced; the first group comprised extracted DNA from130 MDR strains from patients diagnosed previously with TB, and with a history of resistance to first line antituberculosis drugs; the second group comprised 86, DNA extracts directly from sputum samples from consecutive newly diagnosed patients, with clinical symptoms of TB, positive in sputum smear microscopy AFB, chest X-ray positive and presumed drug susceptible with no history of resistance to any first line tuberculosis drugs. The most common mutation found in both groups was in line with other studies, i.e. at position 531, TCG>TTG/Ser>Leu, and at 61.3% and 57.1% for MDR and ND groups respectively, at a similar prevalence.

Studies indicate that 96.1% of the rifampicin resistant strains worldwide will have rpoB mutations (so a surrogate marker for MDR) studies [15, 16]. Comparison of the DST results with the RRDR data from the MDR group in this study suggests that only 61.5% of strains carry mutations in this region of the *rpo*B gene so sequencing the RRDR does not correlate with rifampicin resistance as successfully. The reasons for this may be that resistance to rifampicin is conferred by mutations in other parts of the gene or genome or that the DST testing was sub optimal in some way and indicated resistance when none was present, although all patients from this group had a history of resistance to first line anti-tuberculosis drugs. Antimicrobial resistance testing (AST) or Drug resistance testing was established in the 1960s [17] and there is no consensus reference method for MIC determination against which the different methods can be compared to determine common breakpoints.

The main finding of this study however is the high incidence of rifampicin resistance associated mutations, which is often used as a surrogate marker for MDR. At 16.3%, it is considerably higher than the WHO estimate of 3.9% of new cases of multi drug resistant tuberculosis

[1]. This single piece of study needs to be justified with the help of future studies to support a high percentage of rif resistant strains. The cost of rpoB sequencing in this study was approximately $10, the cost of a GeneXpert test in this region. The time taken to carry out rpoB sequencing is slightly longer than GeneXpert testing but requires more complex testing and analysis. GeneXpert testing or rpoB sequencing for detection of MDR TB in sputum samples is much faster than waiting for results of culture and DST. The diagnosis and effective treatment for individuals with MDR needs to be tailored and administered quickly by rapid molecular tests and, as a consequence of that, the control of transmission of MDR tuberculosis will be tightened. Only then will MDR tuberculosis infection and transmission be effectively controlled in high incidence areas such as Pakistan, where it is most needed.

Supporting information

S1 Fig. Alignment of RRDR sequences from this research work.
(DOC)

Author Contributions

Conceptualization: Salma Hameed, Kartyk Moganeradj, Nasir Mahmood.

Data curation: Kartyk Moganeradj, Catherine Arnold.

Formal analysis: Salma Hameed, Kartyk Moganeradj, Catherine Arnold.

Funding acquisition: Nasir Mahmood.

Investigation: Salma Hameed, Kartyk Moganeradj.

Methodology: Salma Hameed, Kartyk Moganeradj, Catherine Arnold.

Resources: Kartyk Moganeradj, Nasir Mahmood.

Supervision: Kartyk Moganeradj, Nasir Mahmood, Timothy D. McHugh, Muhammad Nawaz Chaudhry, Catherine Arnold.

Writing – original draft: Salma Hameed, Kartyk Moganeradj.

Writing – review & editing: Salma Hameed, Kartyk Moganeradj, Nasir Mahmood, Timothy D. McHugh, Muhammad Nawaz Chaudhry, Catherine Arnold.

References

1. WHO. Global tuberculosis Report. 2016.
2. Negi SS, Khan SF, Gupta S, Pasha ST, Khare S, Lal S. Comparison of the conventional diagnostic modalities, bactec culture and polymerase chain reaction test for diagnosis of tuberculosis. *Indian J Med Microbiol*, 2005, 23 (1):29–33. PMID: 15928418
3. Victor TC, Van Helden PD, Warren R. Prediction of drug resistance in *Mycobacterium tuberculosis*: Molecular mechanisms, tools, and applications. *IUBMB Life*. 2002, 53(4–5):231–237. https://doi.org/10.1080/15216540212642 PMID: 12121001
4. WHO. Anti-tuberculosis drug resistance in the world. Report No.2 Prevalence and Trends. 2000.
5. El-Hajj HH, Marras SA, Tyagi S, Kramer FR, Allan D. Detection of rifampicin resistance in *Mycobacterium tuberculosis* in a single tube with molecular beacons. *J Clin Microbiol*. 2001, 39(11):4131–7. https://doi.org/10.1128/JCM.39.11.4131-4137.2001 PMID: 11682541
6. Mokrousov I, Otten T, Vyshnevskiy B, Narvskaya O. Allele specific rpoB PCR assays for detection of rifampicin resistant *Mycobacterium tuberculosis* in sputum smears. *Antimicrob Agents Chemother*., 2003, 47(7):2231–5. https://doi.org/10.1128/AAC.47.7.2231-2235.2003 PMID: 12821473
7. Kim BJ, Lee KH, Park BN, Kim SJ, Park EM, Park YG, Bai GH, Kim SJ, Kook YH. Detection of rifampicin resistant *Mycobacterium tuberculosis* in sputa by nested PCR linked single strand conformation

polymorphism and DNA sequencing. J Clin Microbiol., 2001, 39(7): 2610–7. https://doi.org/10.1128/JCM.39.7.2610-2617.2001 PMID: 11427578

8. Liu YP, Behr MA, Small PM, Kurn N. Genotype determination of *Mycobacterium tuberculosis* antibiotic resistance using a novel mutation detection method, the branch migration inhibition *M.tuberculosis* antibiotic resistance test. J Clin Microbiol., 2000, 38(10): 3656–62. PMID: 11015379
9. Yang ZT, Durmaz R, Yang D, Gunal S, Zhang L, Foxman B, Sanic A, Marrs CF Simultaneous detection of Isoniazid, Rifampinn and Ethambutol resistance of *Mycobacterium tuberculosis* by single Mutiplex Allele Specific Polymerase Chain Reaction(PCR) assay. Diagn Microbiol Infect Dis., 2005, 53(3): 201–8. https://doi.org/10.1016/j.diagmicrobio.2005.06.007 PMID: 16243477
10. Garibyan L, Avashia N. Polymerase Chain Reaction. Journal of Investigative Dermatology. 2013, 133: 1–4.
11. Bradley P, Gordon NC, Walker TM, Dunn L, Heys S, Huang B, et al. Rapid antibiotic-resistance predictions from genome sequence data for *Staphylococcus aureus* and *Mycobacterium tuberculosis*. *Nature Communications* 2015. 6: 10063. https://doi.org/10.1038/ncomms10063 PMID: 26686880
12. WHO. Global tuberculosis control country profile. World Health Organization Report. 2014.
13. Arnold C, Westland L, Mowat G, Underwood A, Magee J, Gharbia S.005. Single-nucleotide polymorphism-based differentiation and drug resistance detection in *Mycobacterium tuberculosis* from isolates or directly from sputum. *Clinical microbiology and infection*: *the official publication of the European Society of Clinical Microbiology and Infectious Diseases*.; 2005, 11(2):122–30.
14. World Medical Association (WMA). Declaration of Helsinki. Ethical Principles for Medical Research Involving Human Subjects. Jahrb für Wiss und Ethik. 2009; 14. https://doi.org/10.1515/9783110208856.233
15. Mboowa G, Namaganda C, Ssengooba W. Rifampicin resistance mutations in the 81 bp RRDR of *rpo*B gene in *Mycobacterium tuberculosis* clinical isolates using Xpert®MTB/RIF in Kampala, Uganda: a retrospective study. *BMC Infectious Diseases* 2014, 14: 481. https://doi.org/10.1186/1471-2334-14-481 PMID: 25190040
16. Ramaswamy S, Musser JM: Molecular genetic basis of antimicrobial agent resistance in *Mycobacterium tuberculosis*. Tuberc Lung Dis. 1998, 79: 3–29. https://doi.org/10.1054/tuld.1998.0002 PMID: 10645439
17. Schon T, Miotto P, Koser C, Viveiros M, Boettger E, Cambau E. *Mycobacterium tuberculosis* drug resistance testing: challenges, recent developments and perspectives. *Clinical Microbiology and Infection*. 2017, 23(3):154–160. https://doi.org/10.1016/j.cmi.2016.10.022 PMID: 27810467